BIOLOGICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

Journal of Educational Research

OFFICIAL ORGAN OF THE
EDUCATIONAL RESEARCH ASSOCIATION

B. R. BUCKINGHAM, EDITOR

ASSOCIATE EDITORS

E. J. ASHBAUGH W. W. CHARTERS GEORGE D. STRAYER
GUY M. WHIPPLE S. A. COURTIS LEWIS M. TERMAN
FRANK N. FREEMAN

VOL. VII MARCH, 1923 No. 3

CONTENTS

		Page
SOME LIMITATIONS OF EDUCATIONAL TESTS	V. A. C. Henmon	161
THE CONSTRUCTION AND INTERPRETATION OF CORRELATION TABLES	E. L. Thorndike	199
AN EXPERIMENT TO DETERMINE THE EFFECTIVENESS OF PRACTICE TESTS IN TEACHING BEGINNING READING	Nila Banton Smith	213
INTELLIGENCE LEVELS AMONG STATE NORMAL SCHOOL GRADUATES	Frederick L. Whitney	229
SPELLING AGE COMPUTED FROM THE SCORE ON FIFTY PERCENT LISTS	Walter E. Morgan	236
A CONTROLLED EXPERIMENT TO DETERMINE THE EXTENT TO WHICH LATIN CAN FUNCTION IN THE SPELLING OF ENGLISH WORDS	Warren W. Coxe	244
EDITORIALS		248
REVIEWS AND ABSTRACTS		251
NEWS ITEMS AND COMMUNICATIONS		257
EDUCATIONAL RESEARCH ASSOCIATION		273

Entered as second-class matter at the Post Office at Bloomington, Illinois under the act of March 3, 1879.

Journal of Educational Research

Copyright 1923, by
Public School Publishing Co.

Official Organ of the Educational Research Association

Editor-in-Chief, B. R. Buckingham
Director of Bureau of Educational Research, Ohio State University, Columbus, Ohio

---

Associate Editors

E. J. ASHBAUGH, *Ohio State University, Columbus, Ohio.*

W. W. CHARTERS, *Carnegie Inst. Tech., Pittsburgh.*

S. A. COURTIS, *City Schools, Detroit, Mich.*

FRANK N. FREEMAN, *Ex-Officio, Secretary Educational Research Association.*

GEORGE D. STRAYER, *Columbia University, New York City.*

LEWIS M. TERMAN, *Leland Stanford Junior University, Stanford University, California*

GUY M. WHIPPLE, *University of Michigan, Ann Arbor, Michigan.*

Contributing Editors

VIRGIL E. DICKSON, *Director, Bureau of Research and Guidance, Public Schools, Oakland and Berkeley, California.*

CHARLES FORDYCE, *Director of Bureau of Educational Measurements, University of Nebraska, Lincoln, Nebr.*

M. R. HAGGERTY, *Dean of the College of Education, University of Minnesota, Minneapolis.*

V. A. C. HENMON, *Director, School of Education, University of Wisconsin, Madison.*

ARTHUR W. KALLOM, *Assistant Director, Department of Educational Investigation and Measurement, Public Schools, Boston, Mass.*

F. J. KELLY, *Dean of University Administration, University of Kansas, Lawrence, Kans.*

ROBT. H. LANE, *Assistant Superintendent of Schools Los Angeles, Cal.*

WILLIAM A. McCALL, *Assistant Professor of Education, Teachers' College, Columbia University.*

GEORGE W. MELCHER, *Director, Bureau of Research and Efficiency, Public Schools, Kansas City, Missouri.*

EUGENE A. NIFENECKER, *Assistant Director, Bureau of Reference, Research and Statistics, Public Schools, New York City.*

JOSEPH P. O'HERN, *Assistant Superintendent of Schools and Director of Efficiency Bureau, Public Schools, Rochester, New York.*

LEON O. SMITH, *Assistant Superintendent (in charge of Educational Research) Public Schools, Omaha, Nebraska.*

W. W. THEISEN, *Director, Division of Reference and Research, Public Schools, Cleveland, Ohio.*

CLIFFORD WOODY, *Director Bureau of Tests and Measurements, University of Michigan, Ann Arbor, Mich.*

The JOURNAL OF EDUCATIONAL RESEARCH is published at Bloomington, Illinois, monthly, except July and August by the Public School Publishing Co. The subscription price is $4.00; $4.20 in Canada or Mexico; and $4.50 in other foreign countries.

The editorial office is at Columbus, Ohio. Address all communications on editorial matters to B. R. Buckingham, Director of Bureau of Educational Research, The Ohio State University, Columbus, Ohio.

Subscriptions and all communications about reprints or other business matters should be addressed to the Public School Publishing Company at Bloomington, Illinois.

JOURNAL *of* EDUCATIONAL RESEARCH

Volume VII MARCH, 1923 Number 3

SOME LIMITATIONS OF EDUCATIONAL TESTS

V. A. C. HENMON
University of Wisconsin

Standardized educational tests have become so well established and so numerous that it is difficult for even the professional student of education to keep up with them. There are two things about them that are disturbing and indicate how far we must go to make them entirely satisfactory instruments of measurement: (1) the very considerable lack of agreement between scores of the tests and the well-considered judgment of teachers concerning abilities in school subjects indicated by the final marks for a semester or a year's work, and (2) the very considerable lack of agreement between results of different supposedly standard tests in the same subject.

One would surely be justified in expecting that ability or performance in a school subject as measured by one standard test in that subject would show a reasonable agreement with the same ability or performance as measured by another standard test. But this is just what the results do not seem to show. When we give an intelligence test and find a correlation of only 0.50 or 0.60 between it and general scholarship, we hasten to say that success in school work is dependent upon many factors of which intelligence is but one, and that 0.60 is about as high a correlation as we can expect to find. When we give subject-matter tests, on the other hand, we are supposedly measuring classroom products more or less directly, not indirectly, as when we measure native intelligence, and they should surely be expected to show rather close agreement with each other and with teachers' ratings.

Few days go by in which I am not asked to recommend some intelligence test or some standard educational test. This is a common experience to members of departments of education. There is, however, little information available as to the relative merits of different tests and scales, or as to their validity, reliability, and usefulness. Mere opinion, based on respect for an author's standing and workmanship and the popularity of his scale or test, we have wisely learned to distrust.

In order to secure information and a better basis for recommendation, comparative studies have been made of results with different tests in American history, algebra, and reading. In each case, two independent sets of measures were secured.

History Tests

Six American history tests—Bell and McCollum, Barr (Test IV), Buckingham, Starch, Van Wagenen, and Harlan—were given to two senior groups in the high school at Stevens Point, Wisconsin. The first group (1920) numbered sixty; and the second group (1921), sixty-four.[1] The tests were given in the spring after the work of the course for the year, based on Muzzey's "American History" as the text, was completed. The Barr Test IV, and the Van Wagenen Scales were not given to the 1920 classes, but added to the other four in the 1921 tests.

The tests are clearly of two types: those that measure information concerning the facts of history, and those that measure thought or historical judgment. The Bell-McCollum, Starch, and Harlan tests are informational tests only; Barr's Test IV, which was the only one of his tests used, is a thought or judgment test, while Buckingham and Van Wagenen have tests of both types. The scores and correlations were worked out separately in the last two tests. As criteria, against which the scores might be checked, the average of the two semester's marks in history, and the average of the grades assigned in all other subjects, and a composite score based on all the tests, were used.

[1] From an unpublished Master's thesis by Paul M. Vincent.

The correlations with the teacher's marks and intercorrelations for the 1920 group appear in Table I and those for the 1921 group in Table II. The correlations were calculated by the

TABLE I. COEFFICIENTS OF CORRELATION FOR SIXTY HIGH-SCHOOL SENIORS IN VARIOUS HISTORY TESTS (1920)

	History Grades	Average of Other Grades	Starch	Buckingham Thought	Buckingham Information	Harlan	Bell and McCollum
History Grades	.....						
Average of Other Grades	.56	.....					
Starch	.57	.32	.....				
Buckingham Thought	.30	.36	.19	.....			
Buckingham Information	.49	.35	.57	.40	.....		
Harlan	.72	.43	.68	.23	.66	.....	
Bell and McCollum	.65	.38	.70	.22	.65	.82	.....
Average	.50	.40	.50	.32	.52	.59	.57

TABLE II. COEFFICIENTS OF CORRELATION FOR SIXTY-FOUR HIGH-SCHOOL SENIORS IN VARIOUS HISTORY TESTS (1921)

	History Grades	Average of Other Grades	Starch	Buckingham Thought	Buckingham Information	Harlan	Bell and McCollum	VanWagenen Thought	VanWagenen Information	Barr
History Grades	.....									
Average of Other Grades	.72	.....								
Starch	.44	.35	.....							
Buckingham Thought	.34	.21	.23	.....						
Buckingham Information	.45	.42	.27	.23	.....					
Harlan	.60	.52	.51	.18	.31	.....				
Bell and McCollum	.65	.56	.71	.16	.36	.72	.....			
VanWagenen Thought	.42	.37	.50	.26	.39	.52	.52	.....		
VanWagenen Information	.58	.56	.65	.19	.24	.76	.64	.58	.....	
Barr	.12	.22	-.08	.07	.19	.02	.07	.23	.04	.....
Average	.48	.44	.39	.21	.32	.46	.49	.42	.47	.09

product-moment formula. The tables have not been burdened by the addition to each coefficient of its probable error. Anyone can determine it by the usual formula, $P.E._r = .6745 \frac{1 - r^2}{N}$. For a coefficient of 0.50 it is in these tables approximately 0.06 which indicates the limits within which the coefficients are reasonably reliable.

These tables show both the considerable lack of agreement between the teacher's final marks and scores in the tests, and also the lack of agreement between the tests themselves. The correlations between the teacher's marks and the scores in the tests range from 0.12 to 0.72 with an average for thirteen coefficients of only 0.48. This is not an impressive agreement, and does not inspire great confidence in the validity of the measuring instruments. It is somewhat surprising and disconcerting to find that we could predict the teacher's final mark in history by taking the average of the pupil's marks in all other subjects, as well as by the scores in any of the standard tests in history. In fact, the prediction would be better than with the majority of the tests. In Group II the correlation between the history grades and the average of all the other grades, 0.72, is higher than the correlation between the teacher's marks and the history tests, except one. One may, to be sure, urge that the teacher's judgment is subject to much error and that the criterion is inadequate. On the other hand, a teacher is not likely to be interested in nor likely to use a forty-minute test, and place reliance upon it, when its results differ greatly from the judgment based on a year's acquaintance with a pupil's work and ability. He will challenge the validity of the test and decline to use it, particularly if he experiments with several different tests and finds such disagreement between them that the results depend a great deal on the test used. The agreement between the different tests for the 124 pupils ranges from —0.08 to 0.76 with an average for 43 coefficients of only 0.40, again an unimpressive agreement, and one that breeds scepticism concerning the reliability of the tests.

The situation is improved somewhat if the coefficients are considered separately for the informational tests and the judgment tests. The average of the coefficients between the informational tests alone and the teacher's marks is 0.57 instead of 0.48, while of four coefficients based on thought questions alone the average is only 0.295. The intercorrelations between the different informational tests range from 0.24 to 0.82 with an average of 0.58 for sixteen coefficients. For the three coefficients of intercorrelation between the thought tests, the average is only 0.19. It is evident that the teacher's judgment of the pupil's ability in this case, at least, is based upon his usable knowledge of essential historical facts. This is a relatively definite objective thing, and it is likely that the great majority of teachers use it in judging pupil's knowledge and ability in history. Tests of the informational type are likely to be rated most highly by teachers, therefore, and to be most widely used. On this basis of selection the Harlan and Bell-McCollum tests would rank distinctly ahead of the others in practicability.

If, instead of using the teacher's marks, one uses a composite score by combining the scores in all the tests equated to give each test equal weight, the correlations are those given in Table III.

TABLE III. CORRELATIONS WITH COMPOSITE OF ALL TESTS AS CRITERION

	1920 Group	1921 Group
Starch	.87	.83
Bell-McCollum	.88	.78
Harlan	.86	.70
VanWagenen Thought	...	.74
Buckingham Information	.85	.55
VanWagenen Information	...	.70
Buckingham Thought	.48	.45
Barr	...	.29
Teacher's marks	.61	.63

It is noteworthy that the correlations between the composite and teacher's marks are not as high as with several of the tests taken singly. Even if the information tests alone are pooled into a composite the correlations with teacher's marks is 0.66 which is no higher than with one of the tests alone.

By this method of evaluation the tests occupy about the same relative positions except that the Starch test ranks considerably higher. The Van Wagenen thought test takes a better position than his information test.

Algebra Tests

Four algebra tests—Monroe, Rugg-Clark, Hotz, and Stromquist—were given to 100 first-year pupils, fifty boys and fifty girls, in an Omaha high school at the close of the school year.[2] The Coleman test was given to some of the pupils, but the results are not reported here. The Douglas tests were not available when these experiments were made. In order to secure two sets of coefficients the pupils were divided into two groups, A and B, fifty in each group. As criteria against which to check the test scores the teacher's marks for the year's work, and the scores in Army Alpha, which happened to be available, were used. The Pearson correlations and intercorrelations were determined for attempts, rights, and also for an accuracy score based on the relation between attempts and rights. The coefficients based on rights are reported in Table IV.

TABLE IV. COEFFICIENTS OF CORRELATION FOR ONE HUNDRED FIRST YEAR HIGH-SCHOOL PUPILS IN VARIOUS ALGEBRA TESTS

		Algebra Grades	Army Alpha	Monroe	Rugg-Clark	Hotz
Army Alpha	A.	.37				
	B.	.15				
Monroe	A.	.50	.26			
	B.	.45	.29			
Rugg-Clark	A.	.67	.30	.67		
	B.	.39	.30	.77		
Hotz	A.	.54	.32	.31	.58	
	B.	.52	.34	.36	.60	
Stromquist	A.	.72	.51	.53	.78	.74
	B.	.60	.23	.61	.80	.70

[2] From an unpublished Master's thesis by Maybel Ward Burns.

The correlations with the teacher's marks for the different tests ranged from 0.39 to 0.72 with an average for the eight coefficients of 0.55. The incorrelations between the different tests range from 0.31 to 0.80 with an average for the twelve coefficients of 0.62. The results show substantially the same situation as in the history tests. Unfortunately, the average of the marks in other subjects was not obtained, but from numerous other studies of such relationships, it may be inferred that the correlations between the algebra marks and the average of marks in other subjects would be in the neighborhood of 0.50 or 0.60. In other words, prediction of the final mark in algebra could have been made almost as well by the average of all other marks as by any algebra test.

If agreement with the teacher's marks is used as a basis for selection, there are no marked differences, as in history, in the relative values of the tests, though the least well known—the Stromquist test—seems somewhat superior to the others in this case.

If a composite score based on the scores in the four tests equated to give equal weight to each test is used as a criterion, the correlations are as in Table V.

TABLE V. CORRELATIONS WITH THE COMPOSITE OF ALL TESTS AS CRITERION

	Group A	Group B
Rugg-Clark	.92	.94
Stromquist	.92	.91
Monroe	.75	.77
Hotz	.75	.72
Teacher's Marks	.73	.53

As in the history tests the agreement with the teacher's judgment is no closer than with several of the tests taken alone. From this point of view the validity and reliability could not be improved by using all the tests. It might be that a combination of two of the tests could be found which would yield a higher correlation, but in view of the relatively low correlation between the composite and the teacher's marks, it scarcely seemed worth

while to exhaust the permutations and combinations to discover such a team of tests.

With the composite as a criterion the Rugg-Clark and Stromquist tests yield notably higher correlations than the other two.

Reading Tests

The following reading tests were given to two groups of pupils, twenty-seven in a fifth grade, and thirty-one in a sixth grade:[3]

1. Courtis Silent Reading Test, No. 2.
2. Fordyce Scale for Measuring Achievement in Reading, Tests I and II.
3. Kelly-Kansas Silent Reading Test, Tests I and II.
4. Monroe Standardized Silent Reading Test, Form I, Tests II and III.
5. Brown Reading Test.
6. Starch Reading Test, No. 5 and 6.
7. Thorndike Scale Alpha 2, Part II.
8. Burgess Russell Sage Foundation Test.

The tests were all given with strict conformity to directions, within the period of one week, during the last month of the

TABLE VI. COEFFICIENTS OF CORRELATION FOR TWENTY-SEVEN FIFTH-GRADE PUPILS AND THIRTY-ONE SIXTH-GRADE PUPILS IN VARIOUS READING TESTS

FIFTH-GRADE—SPEED

	Courtis	Fordyce	Gray	Brown
Courtis	..			
Fordyce	.75	..		
Gray	.62	.66	..	
Brown	.79	.73	.52	..
Starch	.88	.73	.43	.65

SIXTH-GRADE—SPEED

	Fordyce	Gray	Brown	Starch
Fordyce	..			
Gray	.77	..		
Brown	.53	.63	..	
Starch	.63	.57	.42	..
Monroe	.31	.33	.59	.44

[3] From an unpublished Master's thesis by Olaf K. Thollehaug.

TABLE VII. COEFFICIENTS OF CORRELATION FOR FIFTH- AND SIXTH-GRADE PUPILS IN VARIOUS READING TESTS

FIFTH-GRADE—COMPREHENSION

	Courtis	Fordyce	Gray	Starch	Burgess	Kansas
Courtis	..					
Fordyce	.42	..				
Gray	.16	.58	..			
Starch	.40	.14	.25	..		
Burgess	.41	.02	.03	.33	..	
Kansas	.59	.14	.18	.66	.66	..
Thorndike	.56	.21	.29	.50	.59	.77

SIXTH-GRADE—COMPREHENSION

	Fordyce	Gray	Starch	Monroe	Burgess
Fordyce	..				
Gray	.33	..			
Starch	.41	.25	..		
Monroe	.54	.61	.47	..	
Burgess	.49	.50	.38	.67	..
Thorndike	.35	.39	.08	.58	.40

school year. No two tests were given during the same half day. The papers were all scored in accordance with the directions of authors, except the Thorndike test, in which case the papers were scored on a basis of number of questions right in order to rank the pupils on a finer scale. The correlations were worked out by the Spearman rank formula, and appear in Tables VI and VII.

The correlations show a disappointing lack of agreement, both in speed and in comprehension. In speed the coefficients range from 0.38 to 0.88 with an average for twenty coefficients of 0.60. This is higher than the correlations for comprehension which range from 0.02 to 0.77 with an average of 0.40. When measures of speed of reading show a coefficient of resemblance or agreement of only 0.6 and in comprehension of only 0.4 we become suspicious of the validity and reliability of the tests as practical instruments. Reading ability is defined in very different ways and the tests vary considerably in character.[4] One might

[4] Courtis, S. A. *The Gary public schools: Measurement of classroom products.* New York: General Education Board, 1918. 532 pp.

expect a closer agreement between tests similar in general make-up. Thus a closer agreement might be looked for between the Starch, Brown, Gray, and Fordyce tests. These tests give greater emphasis to reproduction of thought and immediate memory. The Kansas, Monroe, Burgess, and Thorndike tests measure the ability to understand and to follow directions, and they make a greater demand upon judgment. While the results tend to support the expectation, there are, however, no marked differences. For the purpose of comparing the results of closely similar types, the Brown and Starch tests were given and scored in the same manner. The correlation for the fifth-grade group was 0.61 and for the sixth-grade group was only 0.34.

Results similar to these have already been reported for reading tests. Richards and Davidson[5] report a correlation of 0.25 between the Kansas and Starch tests. Courtis,[6] speaking of reading tests in the Gary survey, says, "One would seem to have a right to expect that the reading abilities of the children as revealed by one test will show some agreement with the reading abilities revealed by the next test, since both are tests of reading ability. This, however, is precisely what the results do not show." Starch[7] found a closer agreement, the correlations between the Starch and Kansas tests for fifth- and eighth-grade pupils being 0.56 and 0.50. Franzen,[8] in this JOURNAL, reports correlations for three reading tests with one class of 40 children as follows: Monroe-Thorndike-McCall, 0.26; Burgess-Thorndike-McCall, 0.19; and Monroe-Burgess, 0.48. He points out that "it is obvious from the data at hand that we need to investigate our reading tests."

In order to get some measure of the relative values of the various tests, two criteria were employed, one consisting of the

---

[5] Richards, A. M., and Davidson, P. E., "Correlations of single measures of some representative reading tests," *School and Society*, 4:375-77, September 2, 1916.

[6] Courtis, S. A., *Op. cit.*, p. 265.

[7] Starch, D. "Reliability of reading tests," *School and Society*, 8:86-90, July 20, 1918.

[8] Franzen, Raymond. "Attempts at test validation," *Journal of Educational Research*, 6: 145-58, September, 1922.

average of all the tests combined with the teachers grades, the average of the tests counting one-half and the teacher's grades one-half, the other being the teacher's ratings of the pupils in reading ability. In calculating these coefficients the average of the rank in speed and rank in comprehension was used for test scores. The results appear in Tables VIII and IX.

TABLE VIII. CORRELATION BETWEEN READING TESTS AND COMPOSITE SCORE OF ALL TESTS AND BETWEEN READING TESTS AND TEACHER'S RATINGS

Tests	Correlation between Reading Tests and Composite Score of all Tests			Correlation between Reading Tests and Teacher's Ratings		
	Fifth Grade	Sixth Grade	Average	Fifth Grade	Sixth Grade	Average
Courtis	.83	...	.83	.67	...	.67
Gray	.83	.83	.83	.72	.67	.69
Monroe	...	.79	.79	...	.55	.55
Kansas	.77	...	.77	.60	...	.60
Thorndike	.77	.74	.75	.60	.65	.62
Starch	.73	.76	.75	.50	.65	.57
Fordyce	.69	.80	.75	.59	.52	.56
Burgess	.58	.76	.67	.40	.57	.48

As compared with the history and algebra tests the correlations with teacher's estimates are higher, but are disappointingly low. If the teacher's judgment is taken as a criterion, the tests would rank in relative value as they appear in Table VIII. If the composite criterion is chosen, the ranks would differ very little except for the position of the Monroe and Thorndike tests.

If the teacher's ratings be disregarded and a composite from all tests is used as a criterion, the resulting correlations with each test and with the teacher's estimates of reading ability are as in Table IX. The highly erratic character of these results emphasizes in a striking way the unsatisfactory state of our reading tests.

There is an abstractness about correlation coefficients that makes it difficult to realize their concrete significance. An illustration or two will show the situation in individual cases. In a

TABLE IX. CORRELATIONS WITH COMPOSITES OF ALL TESTS

Reading Tests	Speed		Reading Tests	Comprehension	
	Fifth Grade	Sixth Grade		Fifth Grade	Sixth Grade
Courtis	.95	...	Kansas	.85	...
Fordyce	.83	.84	Monroe	...	.81
Brown	.87	.79	Burgess	.68	.80
Starch	.86	.75	Courtis	.70	...
Gray	.70	.86	Starch	.73	.60
Monroe	...	.61	Thorndike	.76	.56
Teacher's Ratings in Reading Ability	.74	.37	Gray	.52	.75
			Brown	.66	.60
			Fordyce	.42	.68
			Teacher's Ratings (Reading)	.66	.55

group of 27 pupils one of the reading tests ranked a pupil second in comprehension; another, sixteenth; and a third, twenty-sixth. In the first case he is near the top, in the second he is an average performer, and in the third he is near the foot of the class. Even in speed, in a group of 31 pupils, one pupil is rated second and third by two tests, twenty-eighth by another test, and thirteenth and fifteenth by two others. Reading ages or achievement ages for such pupils will vary enormously and can have little significance. That such cases can occur at all is a serious matter, but when they occur not only occasionally, but with considerable frequency, the matter becomes still more serious.

Discussion

The results of each set of experiments, and they are in fundamental agreement, are at best discouraging. Thorndike has pointed out that if a test or a team of tests yields a correlation with some accepted criterion of 0.80 and an attempt is made by the test or tests to assign individuals to their proper tenths, we should be right only twenty-three times out of a hundred. Jordan[9] has recently extended the calculations and has shown that with a correlation of 0.70 if we wished to assign individuals to their proper fifths, which would agree with the ordinary five-

[9] Jordan, A. M. "Correlations of four intelligence tests with grades," *Journal of Educational Psychology*, 13:429, October, 1922.

point marking system, we would be right only forty times out of a hundred. If the assignment is made by thirds, or into A, B, and C sections, we would be right only fifty-five times out of a hundred. If that is the case, our educational tests are clearly subject to very marked limitations as practical measuring instruments. It is likewise evident that when the less valuable tests have been weeded out, the remainder will need revision, reconstruction, and supplementation.

When educational tests were first developed, it was thought and frequently urged, that they would be most useful in surveys and in comparing schools and classes but of doubtful value for the study of individual pupils. It was also not infrequently pointed out that the tests were measures of performance in certain specified activities under standardized conditions and could not necessarily be regarded as measures of ability in school subjects. This has been forgotten recently, however, in the elaborate calculations of achievement ages and achievement quotients for individual pupils. These are then taken as measures of ability in school subjects as well as measures of individuals. The writer has no lack of faith in the ultimate development of educational tests, but in the light of such data as are available, he is skeptical of the validity and the reliability of achievement or accomplishment quotients in history, algebra, and reading, and is doubtful whether we can legitimately and safely claim to measure educational products with sufficient accuracy to go much beyond the comparison of schools and classes.

We have been too much influenced by motives of economy in time, money, and energy. In using intelligence tests, we have found a thirty-minute test inadequate except for rough classification in a very heterogeneous group and most students of intelligence tests would probably agree with Thorndike that two hundred minutes of testing is necessary to secure a valid and reliable result with older pupils. Just so, in measuring educational products we must contemplate a more comprehensive and time-consuming program. It is noteworthy that the correlations between a teacher's estimate and the best of the intelligence

tests are higher than in the case of educational tests, and also that the intercorrelations between the different intelligence tests are much higher than between different educational tests in the same subject. It appears almost as if we have been taking both the validity and the reliability of subject-matter tests for granted.

THE CONSTRUCTION AND INTERPRETATION OF CORRELATION TABLES[1]

EDWARD L. THORNDIKE
Institute of Educational Research
Teachers College, Columbia University

It is the purpose of this article to explain and illustrate a method of making correlation tables from given hypotheses about the causes which produce the correlation. Making and inspecting such tables will give an insight into the meaning of correlation tables, diagrams, and coefficients which is limited now to a very few experts in this field. It also results in a number of intrinsically valuable observations. The procedure to be shown also makes it easy to construct a "theoretical" correlation table representing a given hypothesis, and to compare observed correlation therewith, as we construct a theoretical frequency curve and compare an observed set of frequencies with it.

My procedure is almost childish in its mathematics, but it seems to me to be more instructive to most students of psychology and mental measurements than a generalized treatment with calculus would be. I shall be gratified if such a generalization is made by some one possessed of the mathematical technique which I lack. Even if this is done, the present treatment will probably be useful as an introduction to the generalized treatment, and for non-mathematical workers will take the place of it.

Let A and B be two quantities, like height and arm stretch or ability with words and ability with numbers, which vary in magnitude among individuals and are positively, but not perfectly, correlated in the sense that the amount of A which an individual possesses corresponds somewhat to the amount of B which he possesses.

For A and B to be positively but not perfectly correlated means that A and B are produced by certain causes some of which are identical for A and B, whereas others act only on A

[1] This investigation is one of a series made possible by a grant from the Carnegie Corporation.

and others only on B. Call these the AB causes, the A causes, and the B causes, respectively.

Consider a population of 320,000 (taken arbitrarily at this number) and consider the correlations that will result according to certain conditions with respect to the AB, A, and B causes.

Assume that the amounts of A, if only the AB causes acted, would be as shown below:

Amount of A	Frequency in Percents
K	0.5
K+1	1.5
K+2	4.5
K+3	9.0
K+4	15.0
K+5	19.5
K+6	19.5
K+7	15.0
K+8	9.0
K+9	4.5
K+10	1.5
K+11	0.5

(Any form of distribution could be used as well without impairing the later argument. This, which is closely the so-called "normal" distribution, is taken as a form to which approximations are often encountered in mental and social measurements.)

Call the amount of B produced by the AB causes that produce K of A, C. Call the amount of B produced by the AB causes that produce K+1 of A, C+1. (This might be called C+2 or C+⅛ or C+0.7432, or any number without impairing the later argument in any way, but the computations would be much more laborious to make and to follow.)

Case I.—If only the AB causes act, every individual having K of A will have C of B, every individual having K+1 of A will have C+1 of B, and so on. The correlation will of course be perfect and the correlation table will be as shown in Table I. A break below perfect correlation occurs only when the causes influencing the variables A and B are not identical.

TABLE I. CORRELATIONS, ONLY THE AB CAUSES OPERATIVE

	K	K+1	K+2	K+3	K+4	K+5	K+6	K+7	K+8	K+9	K+10	K+11
C	1600											
C+1		4800										
C+2			14400									
C+3				28800								
C+4					48000							
C+5						62400						
C+6							62400					
C+7								48000				
C+8									28800			
C+9										14400		
C+10											4800	
C+11												1600

Case II.—Suppose now that the AB causes act and also certain B causes, but no A causes. Suppose that the B causes are such as have in general this result:

In 2½ cases out of 100 they lower a person by 4
In 5 cases out of 100 they lower a person by 3
In 12½ cases out of 100 they lower a person by 2
In 17½ cases out of 100 they lower a person by 1
In 25 cases out of 100 they lower a person by 0
In 17½ cases out of 100 they raise a person by 1
In 12½ cases out of 100 they raise a person by 2
In 5 cases out of 100 they raise a person by 3
In 2½ cases out of 100 they raise a person by 4

(Any arrangement of resulting effects from the B causes could be used without impairing the later argument. This arrangement is taken as a common case where the causes act as often to raise as to lower B.)

Then the 1600 individuals who would be K in A by the action of the AB causes will still all be just K, but in B 40 of them will be, not C, but C—4; 80 of them will be, not C, but C—3; 200 of them will be C—2; 280 of them will be C—1; 400 of them will be C; and so on. The 4800 individuals who would

be K+1 in A by the action of the AB causes will still all be K+1, but in B 120 of them will be C—3, 240 of them will be C—2, and so on. The correlation table will be as shown in Table II. Table II differs from Table I in two respects. First, the variability of the population in B has been notably increased by the addition of the B causes to the AB causes. Second, whereas, with perfect correlation, individuals who were C in B averaged K in A, they now "regress" toward the general central tendency of A, averaging about K+2.4. Similarly those individuals who are C+1 in B do not now center at K+1 but nearly at K+3.

Case III.—Suppose now that the AB causes act as hitherto, and the B causes as in the previous case, and that there also act certain A causes such as have in general this result:

In 7½ cases out of 100 or 3 in 40 they lower a person's A by 2
In 22½ cases out of 100 or 9 in 40 they lower a person's A by 1
In 40 cases out of 100 or 16 in 40 they lower a person's A by 0
In 22½ cases out of 100 or 9 in 40 they raise a person's A by 1
In 7½ cases out of 100 or 3 in 40 they raise a person's A by 2

We now have for any 1600 individuals who would, by the AB causes alone, be K and C or K+1 and C+1, or K+2 and C+2, a scattering as follows: 120 of them will be lowered 2 in A; of these 3 will be lowered 4 in B, 6 will be lowered 3 in B, 15 will be lowered 2 in B, 21 will be lowered 1 in B, 30 will be lowered 0 in B, 21 will be raised 1 in B, and so on. 360 of them will be lowered 1 in A; of these 9 will be lowered 4 in B, 18 will be lowered 3 in B, 45 will be lowered 2 in B, 63 will be lowered 1 in B, and so on.

If, for example, we take the 4800 individuals who, by the AB causes alone, would all be K+1 in A and C+1 in B, they will scatter as shown in Table III. If we take the 320,000 individuals that would be as in Table I by the AB causes alone they will produce Table IV. The variability of both A and B are increased by the addition of the A and B causes to the AB causes, and there is now "regression" both toward the central tendency in A and toward the central tendency in B.

	K	K+1	K+2	K+3	K+4	K+5	K+6	K+7	K+8	K+9	K+10	K+11	Total
C−4	40												40
C−3	80	120											200
C−2	200	240	360										800
C−1	280	600	720	720									2320
C	400	840	1800	1440	1200								5680
C+1	280	1200	2520	3600	2400	1560							11560
C+2	200	840	3600	5040	6000	3120	1560						20360
C+3	80	600	2520	7200	8400	7800	3120	1200					30920
C+4	40	240	1800	5040	12000	10920	7800	2400	720				40960
C+5		120	720	3600	8400	15600	10920	6000	1440	360			47160
C+6			360	1440	6000	10920	15600	8400	3600	720	120		47160
C+7				720	2400	7800	10920	12000	5040	1800	240	40	40960
C+8					1200	3120	7800	8400	7200	2520	600	80	30920
C+9						1560	3120	6000	5040	3600	840	200	20360
C+10							1560	2400	3600	2520	1200	280	11560
C+11								1200	1440	1800	840	400	5680
C+12									720	720	600	280	2320
C+13										360	240	200	800
C+14											120	80	200
C+15												40	40
Total	1600	4800	14400	28800	48000	62400	62400	48000	28800	14400	4800	1600	320000

be K+1 in A by the action of the AB causes will still all be K+1, but in B 120 of them will be C—3, 240 of them will be C—2, and so on. The correlation table will be as shown in Table II. Table II differs from Table I in two respects. First, the variability of the population in B has been notably increased by the addition of the B causes to the AB causes. Second,

TIGHT BOUN

will scatter as shown in Table III. If we take the 320,000 individuals that would be as in Table I by the AB causes alone they will produce Table IV. The variability of both A and B are increased by the addition of the A and B causes to the AB causes, and there is now "regression" both toward the central tendency in A and toward the central tendency in B.

	K																							
C−4	40																							
C−3	80																							
C−2	200																							
C−1	280																							
C	400																							
C+1	280																							
C+2	200																							
C+3	80																							
C+4	40																							
C+5																								
C+6																								
C+7																								
C+8																								
C+9																								
C+10																								
C+11																								
C+12																								
C+13																								
C+14																								
C+15																								
Total	1600	4800	14400	28800	48000	62400	62400	48000	28800	14400	4800	1600	320000											

	K																							
C−4	40																							
C−3	80																							
C−2	200																							
C−1	280																							
C	400																							
C+1	280																							
C+2	200																							
C+3	80																							
C+4	40																							
C+5																								
C+6																								
C+7																								
C+8																								
C+9																								
C+10																								
C+11																								
C+12																								
C+13																								
C+14																								
C+15																								
Total	1600	4800	14400	28800	48000	62400	62400	48000	28800	14400	4800	1600	320000											

HTLY
UND

TABLE III. ANALYSIS OF THE 4,800 INDIVIDUALS WHO BY AB CAUSES ALONE WOULD BE $k+1$ IN A AND $c+1$ IN B

	K−1	K	K+1	K+2	K+3	Total
C−3	9	27	48	27	9	120
C−2	18	54	96	54	18	240
C−1	45	135	240	135	45	600
C	63	189	336	189	63	840
C+1	90	270	480	270	90	1200
C+2	63	189	336	189	63	840
C+3	45	135	240	135	45	600
C+4	18	54	96	54	18	240
C+5	9	27	48	27	9	120
Total	360	1080	1920	1080	360	4800

Case IV.—Consider now a more complicated arrangement of the A causes and B causes. Let the AB causes be as before. Let the A causes act as in the case of Table IV. Let the B causes now vary in their nature in relation to the magnitude of the B as determined by the AB causes, and in such fashion that when the AB causes alone would produce C of B, the B causes produce effects of —1 three times in 40, 0 thirty-four times in 40, and +1 three times in 40; when the AB causes alone would produce C+1, the B causes produce effects of —1 six times in 40, 0 twenty-eight times in 40, and +1 six times in 40; and so on as shown in Table V.

Table IV and any table of the same sort reveals an inadequacy in the orthodox teaching that the variability of any array of the B's in a "normal" correlation table will be the same as that of any other array of the B's, and will be $\sqrt{1-r^2}$ times the variability of the total population of the B's concerned in the correlation. The arrays of the B's toward one or both of the extremes will on the contrary always have less variability than those toward the center, since their variability is at the extremes only that produced by the B causes, whereas as we

TABLE IV. CORRELATIONS, AB CAUSES, B CAUSES, AND A CAUSES OPERATIVE

K−2	K−1	K	K+1	K+2	K+3	K+4	K+5	K+6	K+7	K+8	K+9	K+10	K+11	K+12	K+13	Total

TABLE IV. CORRELATIONS, AB CAUSES, B CAUSES, AND A CAUSES OPERATIVE

	K−2	K−1	K	K+1	K+2	K+3	K+4	K+5	K+6	K+7	K+8	K+9	K+10	K+11	K+12	K+13	Total
C−4	3	9	16	9	3												40
C−3	6	27	59	66	33	9											200
C−2	15	63	161	222	213	99	27										800
C−1	21	108	301	519	606	495	216	54									2320
C−0	30	153	484	939	1353	1314	939	378	90								5680
C+1	21	153	571	1380	2289	2754	2310	1434	531	117							11560
C+2	15	108	539	1569	3228	4473	4623	3327	1776	585	117						20360
C+3	6	63	356	1365	3399	5907	7158	6342	3903	1767	504	90					30920
C+4	3	27	205	888	2811	5958	9111	9381	7071	3696	1413	342	54				40960
C+5		9	81	480	1755	4671	8553	11307	9966	6378	2826	918	189	27			47160
C+6			27	189	918	2826	6378	9966	11307	8553	4671	1755	480	81	9		47160
C+7				54	342	1413	3696	7071	9381	9111	5958	2811	888	205	27	3	40960
C+8					90	504	1767	3903	6342	7158	5967	3399	1365	356	63	6	30920
C+9						117	585	1776	3327	4623	4473	3228	1569	539	108	15	20360
C+10							117	531	1434	2310	2754	2289	1380	571	153	21	11560
C+11								90	378	939	1314	1353	939	484	153	30	5680
C+12									54	216	495	606	519	301	108	21	2320
C+13										27	99	213	222	161	63	15	800
C+14											9	33	66	59	27	6	200
C+15												3	9	16	9	3	40
Total	120	720	2800	7680	17040	30600	45480	55560	55560	45480	30600	17040	7680	2800	720	120	320000

approach the center the A causes act to commingle in one array individuals who by the AB causes alone would have been kept in distinct arrays. So some of the AB causes of variability combine with the B causes to produce the observed variability of the more central arrays. The same holds for arrays of the A's against any one value of B.

The resulting correlation table is Table VI. This shows that although the distribution of variable A and variable B by the AB causes alone was symmetrical and of the "normal" form, and although the effect of the A causes at each single magnitude of A (as determined by the AB causes only), and of the B causes at each single magnitude of B (as determined by the AB causes only), was symmetrical and of the "normal" form, the increase of the effect of the B causes in relation to the AB magnitude has produced a skew in the variable B in the whole population, a skew in each array of B's, (save the two extremes) and a bend of the line showing the relation of A to B.

TABLE V. THE EFFECTS PRODUCED BY THE B CAUSES IN RELATION TO MAGNITUDE OF B DUE TO THE AB CAUSES ONLY

Frequencies of Effects of the B Causes	Magnitude of B											
	C	C +1	C +2	C +3	C +4	C +5	C +6	C +7	C +8	C +9	C +10	C +11
−6											1	1
−5										1	1	2
−4								1	1	2	2	2
−3					1	1	2	2	4	3	3	3
−2				2	2	4	5	5	4	5	4	4
−1	3	6	10	10	10	9	8	7	7	6	5	5
0	34	28	20	16	14	12	10	10	8	6	8	6
+1	3	6	10	10	10	9	8	7	7	6	5	5
+2				2	2	4	5	5	4	5	4	4
+3					1	1	2	2	4	3	3	3
+4								1	1	2	2	2
+5										1	1	2
+6											1	1

Case V.—Consider now a case where the AB causes act as before, and where both the A causes and the B causes act as the B causes did in Case IV (Table VI). The resulting correlation table is Table VII. We have skew in both directions, skew in all

TABLE VI

	K−2	K−1	K	K+1	K+2	K+3	K+4	K+5	K+6	K+7	K+8	K+9	K+10	K+11	K+12	K+13	Total
C−1	9	27	48	27	9												120
C	102	360	706	594	264	54											2080
C+1	9	279	1074	2289	2619	1908	1074	378	90								9720
C+2		54	702	2448	4842	5211	3471	1704	531	117							19080
C+3			270	1674	4932	8586	9300	6852	3822	1650	504	90					37680
C+4				540	2880	7713	12084	11871	8133	3957	1503	513	183	54	9		49440
C+5				108	1224	4680	10272	13554	11670	7314	3609	1416	435	97	18	3	54400
C+6					180	1593	5289	10296	11685	8652	4473	1770	573	167	36	6	44720
C+7					90	738	2820	6474	9564	9345	6084	2841	945	248	45	6	39200
C+8						117	936	3009	5793	6690	5301	2907	1137	318	63	9	26280
C+9							234	1152	2976	4398	4131	2595	1140	361	81	12	17080
C+10								180	756	1770	2250	1923	1131	458	117	15	8600
C+11								90	486	1263	1872	1611	915	366	99	18	6720
C+12									54	243	567	717	534	269	81	15	2480
C+13										54	189	381	342	199	63	12	1240
C+14										27	99	207	204	129	45	9	720
C+15											9	33	66	59	27	6	200
C+16											9	33	66	59	27	6	200
C+17												3	9	16	9	3	40
Total	120	720	2800	7680	17040	30600	45480	55560	55560	45480	30600	17040	7680	2800	720	120	320000

TABLE VII

	K −1	K	K +1	K +2	K +3	K +4	K +5	K +6	K +7	K +8	K +9	K +10	K +11	K +12	K +13	K +14	K +15	K +16	K +17	Total
K−1	9	102	9																	120
K	102	1264	606	108																2080
K+1	9	606	3363	2724	1776	780	372	60	30											9720
K+2		108	2724	5667	5436	2991	1428	471	216	39										19080
K+3			1776	5436	9318	8934	6222	2994	1848	756	306	60	30							37680
K+4			780	2991	8934	11964	10725	6474	4164	1956	885	270	192	57	27	15	3	3		49440
K+5			372	1428	6222	10725	12262	9332	6773	3777	2005	725	549	143	49	27	5	5	1	54400
K+6			60	471	2994	6474	9332	9034	7345	4539	2534	928	675	187	80	45	10	10	2	44720
K+7			30	216	1848	4164	6773	7345	7288	5136	3329	1456	1086	307	125	69	13	13	2	39200
K+8				39	756	1956	3777	4539	5136	4017	2844	1431	1134	369	156	87	18	18	3	26280
K+9					306	885	2005	2534	3329	2844	2203	1268	1023	368	169	96	23	23	4	17080
K+10					60	270	725	928	1456	1431	1268	949	768	355	200	117	34	34	5	8600
K+11					30	192	549	675	1086	1134	1023	768	654	297	159	93	27	27	6	6720
K+12						57	143	187	307	369	368	355	297	172	110	66	22	22	5	2480
K+13						27	49	80	125	156	169	200	159	110	79	48	17	17	4	1240
K+14						15	27	45	69	87	96	117	93	66	48	30	12	12	3	720
K+15						3	5	10	13	18	23	34	27	22	17	12	7	7	2	200
K+16						3	5	10	13	18	23	34	27	22	17	12	7	7	2	200
K+17							1	2	2	3	4	5	6	5	4	3	2	2	1	40
Total	120	2080	9720	19080	37680	49440	54400	44720	39200	26280	17080	8000	6720	2480	1240	720	200	200	40	320000

arrays save the extremes, and bend in the relation lines of A to B and of B to A (which are identical here). The bend is shown in Figure 1.

One of the important uses of correlation tables is in the case of "self-correlation" where two imperfect measures of the same thing are correlated, the drop below perfect correlation or the "attenuation" being an index of the imperfection of the measures. In this case the AB causes are those which act to produce perfect measures, the A causes being the causes of "error" in the A measures, and the B causes being the causes of "error" in the B measures. One may therefore construct correlation tables to represent any hypothesis in which one is interested concerning the nature and amount of the errors in A and B. For example, let us assume that A and B are imperfect measures of the same fact; that if this fact were perfectly measured it would be distributed as shown in Table VIII; that the A measures of it are afflicted with an error such as makes a person's A score too small

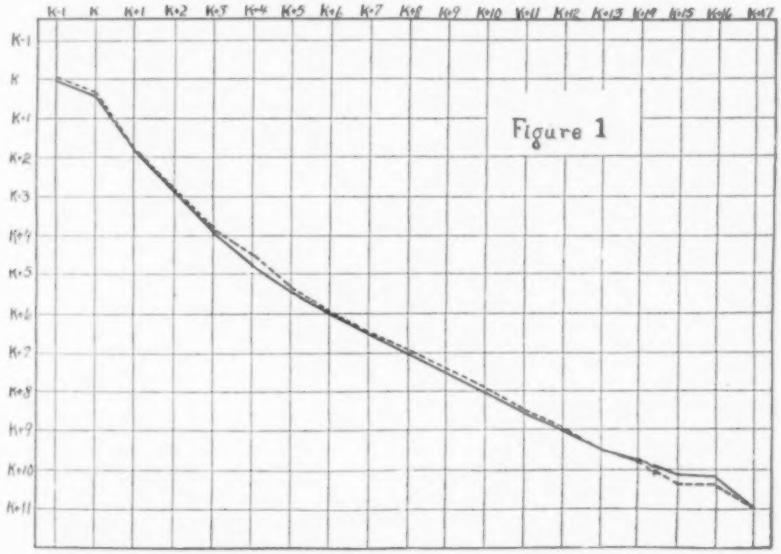


FIGURE 1. THE RELATION LINE OF TABLE VII. THE CONTINUOUS LINE IS FOR THE AVERAGES OF THE ARRAYS, THE DASH LINE IS FOR THE MEDIANS OF THE ARRAYS

or too large as shown in Table IX; and that the B measures of it are afflicted by an error which makes a person's B score too small or too large as shown in Table X. Since these suppositions duplicate in statistical effect the suppositions of Case III above, the correlation table will be exactly Table IV, except that for C—2, C—1, C, C+1, etc., we should read K—2, K—1, K, etc., K+1, etc., since the A and B series are both measures of the same thing.

As a second example, let us assume that A and B are imperfect measures of the same fact; that, if this fact were perfectly measured in the group in question, it would be distributed

TABLE VIII. THE DISTRIBUTION OF THE FACT WHEN PERFECTLY MEASURED

Amount of A	Frequency, in Percents
K	0.5
K+1	1.5
K+2	4.5
K+3	9.0
K+4	15.0
K+5	19.5
K+6	19.5
K+7	15.0
K+8	9.0
K+9	4.5
K+10	1.5
K+11	0.5

TABLE IX. AMOUNT OF ERROR DUE TO THE A CAUSES

In 3 cases in 40 they lower a person's score in A by 2
In 9 cases in 40 they lower a person's score in A by 1
In 16 cases in 40 they lower a person's score in A by 0
In 9 cases in 40 they raise a person's score in A by 1
In 3 cases in 40 they raise a person's score in A by 2

TABLE X. AMOUNT OF ERROR DUE TO THE B CAUSES

In 1 case in 40 they lower a person's score in B by 4
In 2 cases in 40 they lower a person's score in B by 3
In 5 cases in 40 they lower a person's score in B by 2
In 7 cases in 40 they lower a person's score in B by 1
In 10 cases in 40 they lower a person's score in B by 0
In 7 cases in 40 they raise a person's score in B by 1
In 5 cases in 40 they raise a person's score in B by 2
In 2 cases in 40 they raise a person's score in B by 3
In 1 case in 40 they raise a person's score in B by 4

as shown in Table VIII; that the A measures of it are afflicted with an error such as makes a person's score too small or too large as shown in Table V; and that the B measures of it are afflicted by an error of the same nature and magnitude. Since these suppositions duplicate in statistical effect the suppositions of Case V above, the correlation table will be exactly Table VII except that for C—1, C, C+1, C+2, and so on, we should read K—1, K, K+1, K+2, etc.

The admixture of causes of error (A and B causes) to causes of the true facts (AB causes) increases the variability of the distributions, this being the effect discussed by Thorndike[2] and by Kelley.[3] It lowers the correlation, this being the attenuation discovered and fully described by Spearman.[4] It may also change the form of the distribution and change the form of the relation line. Thus if the error increases in relation to the magnitude, the distribution of the imperfectly measured fact will, other things being equal, take on skew (or more skew, if the perfectly measured fact had some in the same direction), all of the arrays save the extremes will take on skew, and the relation line will take on bend (or more bend if it was originally bent in the same direction).[5]

Corrections for the effect of errors of certain type in increasing variability have been worked out by Kelley; corrections for their effects in reducing correlation coefficients have been worked out by Spearman; and corrections for the effect of errors related in magnitude to the magnitude of the fact in adding skewness to distributions and bending relation lines can probably be worked out also. I am, however, convinced that

[2] Thorndike, E. L. *Introduction to the theory of mental and social measurements.* New York: The Science Press, 1913, p. 93.

[3] Kelley, T. L. "The measurement of overlapping," *Journal of Educational Psychology,* 10:458-461, November, 1919.

[4] Spearman, C. "The proof and measurement of association between two things," *American Journal of Psychology,* 15:72-101, January, 1904. "General intelligence objectively determined and measured," *American Journal of Psychology,* 15:201-92, April, 1904.

[5] The skew or bend may of course counterbalance a skew or bend of opposite sense and so give a spurious symmetry for a distribution that with perfect measures would be skewed, or a spurious rectilinearity for a relation line that with perfect measures would be curved.

for most of our work we should try to reduce the magnitude of these errors at the outset, as well as to correct faithfully for such as remain. At the present time more work is being done probably with measures whose self-correlation in the group in question is under 0.80 than with measures whose self-correlation is over 0.80.

Such measures do not allow anything like accurate location of an individual, and require elaborate correction to reach correct ideas about the form of distribution of the group, the nature and amount of overlapping between groups, the amount of correlation, and the form of the relation line. We do not yet even know how to make some of these corrections. It seems, therefore, the part of wisdom to make much more adequate original measures than it has been our custom to make.

AN EXPERIMENT TO DETERMINE THE EFFECTIVENESS OF PRACTICE TESTS IN TEACHING BEGINNING READING

NILA BANTON SMITH
Assistant Supervisor of Primary Grades, Detroit, Michigan

During this period of reconstruction, there probably is no other subject in the school curriculum which is receiving such widespread and concentrated attention as reading. Since it is generally conceded that reading is the most important of the school arts, that it is the master key with which we unlock the gateway to every other subject, it is small wonder that reading has been the target for such universal investigation. These scientific studies, resulting in valuable contributions to the teacher, have been carried on by many competent people in all parts of the country. They have extended over a very long period of time and have covered all stages of reading development. As a result we have accumulated a great body of scientific knowledge in which the process of learning to read has been analyzed into its elements and the laws governing the successful teaching of reading have been clearly formulated.

By combining the most significant of these contributions with a vast amount of experimental data accumulated while attempting to develop practice tests in reading, it was possible to set up certain definite aims or objectives which it seemed highly desirable to embody and combine in a *complete new method* for teaching beginning reading to first-grade children. The most important of these aims were:

1. To develop the child's power to purpose through the use of reading situations.
2. To enable him to use his reading ability as a tool in carrying out his purpose from the time of his first contact with the reading activity.
3. To provide some means whereby the child may largely teach himself in learning to read.

4. To permit each child to progress as rapidly as his individual effort and ability will permit.

It might be said that there were at least a score of other aims which were constantly kept in mind throughout the preparation of the material. But the ones mentioned above, "purposing," "reading as a tool," "self-helpfulness," and "individual progression," were the four corner-stones upon which the framework of the Practice Tests was built.

Description of Material

The materials developed to meet these aims consisted of several different pieces. In the account below each of the pieces needed for one complete lesson is described, the sequence of the steps of procedure is shown, and the use to which each piece of material was put is explained.

1. *Picture Frame.*—The child is first provided with a picture frame on which is printed a background, but in which the important objects are omitted and in their places are printed their names.

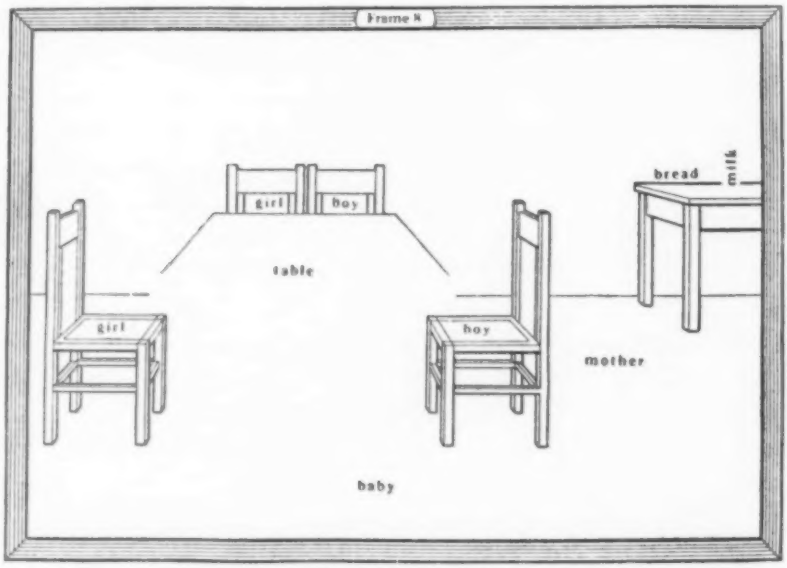


FIGURE 1. PICTURE FRAME

2. *"Cut-outs."*—The child is next given a sheet on which are printed these missing objects which are to be cut out, colored, and pasted on the frame to make the completed picture.

FIGURE 2. "CUT-OUTS"

It is reasonably safe to assume that in every little child there exists an innate love for cutting, coloring, and pasting. In the Practice Test Material this instinctive tendency is taken advantage of in developing the child's purpose to read. When he sees the frame and the cut-outs arrayed before him there is aroused within him the desire to construct the completed picture and he is willing to carry on any activity which will aid him in attaining this end. In this particular case the activity involved is learning to read a set of directions in which he is given instructions for cutting, coloring, and placing the various figures in the frame.

3. Book of Directions.—The directions needed in building up the pictures are given in a little book called the "Book of Directions." The directions for arranging this picture taken from pages 20 and 21 are as follows:

Cut out the milk.
Cut out the bread.
Cut out the ice-cream.
Cut out the candy.
Cut out the cake.
Cut out the baby.
Cut out the mother.

Do not color the table.
Do not color the girls.
Do not color the boys.
Color the candy red.
Do not color the milk.
Color the bread brown.
Do not color the ice-cream.
Color the cake brown.
Do not color the baby.
Do not color the mother.

Paste the girls in the frame
where you see girls.
Paste the boys in the frame
where you see boys.
Paste the table on the frame
where you see table.
Paste the candy on the table
where you see candy.
Paste the ice-cream
on the table
where you see ice-cream.
Paste the cake on the table
where you see cake.
Paste the bread
on the little table
where you see bread.
Paste the milk
on the little table
where you see milk.

The child's motive in reading these directions is analogous to that of the soldier of fortune in the conventional treasure island stories, who learns that gold is buried in a certain area. Some dying man whom he has befriended hands him a code which shows the exact location of the treasure. He cannot find it, however, until he is able to decipher the cryptic directions.

Hence he has a powerful motive for learning to read the mysterious symbols. The new words involved in each new set of directions are symbols which are probably as mysterious to the little child as were the ciphers in the code of the soldier of fortune. How is he going to find out what these new words are so that he may carry out the directions for making his picture? It is not necessary for him to wait until the teacher develops the words and gives the whole class drill upon them. He may find them out for himself just as rapidly as he cares to look them up.

FIGURE 3. THE "DICTIONARY"

4. Dictionary.—As a reference for looking up words, the child is furnished a "Dictionary" in which all new words needed are pictured in such a way that he may find them for himself by matching the unrecognized word with the same word in the dictionary and looking at the picture above it.

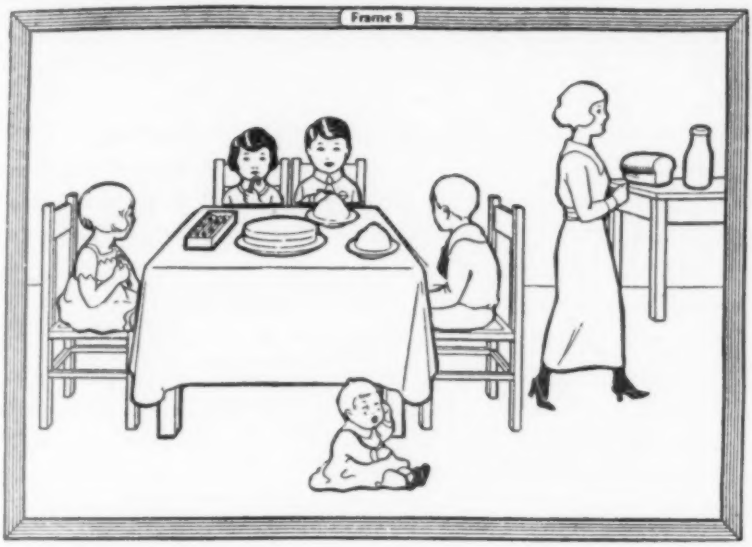
5. Story Sheet.—Upon completing his picture the child is given a loose-leaf sheet on which is printed a small replica of the picture he has made together with a story about the picture which involves all the new words which he has been teaching himself while building the picture.

The small reproduction of the completed picture furnishes the child with a standard whereby he may judge of the success or failure of his work. He compares his product with this one and if the two are exactly alike he knows that he has carried out the directions correctly. In other words he has a means of checking up his own comprehension. If his picture is correct he then studies the story about the picture with reference to his dictionary, if necessary. When he feels that he is prepared, he goes to the teacher and reads it aloud to her. This oral reading of the story is a test of the child's ability to recognize the new words learned during the picture-building activity.

6. Cover for Loose Leaf Sheets.—The child has in his possession a cover called "My Story Book" in which he files away his story sheets as rapidly as he is able to earn them. The number of pages contained in this book at any time is a measure of the child's progress. The pupil himself takes very great interest in watching his story book grow in volume and it also furnishes tangible evidence whereby parents may constantly keep themselves acquainted with the rate and content of their child's progress.

Description of Experiment

This special type of reading was carried out with 180 beginning first-grade children in Detroit schools during a period of five months, September, 1921 to February, 1922. Four schools of different types were selected as the field of experimentation—a platoon school attended by children from the best homes, a



The cake, candy
and ice-cream
are on the big table.
The bread and milk
are on the little table.
The boys are eating
the ice-cream and cake.
The girls are eating candy.
Baby is crying.
He cannot eat
ice-cream and cake.
Mother will give him bread
and milk.

FIGURE 4. STORY SHEET

plattoon school attended by children from middle-class homes, an ordinary type of school attended by children from middle-class homes and an ordinary type of school attended by foreign (Polish and Hungarian) children who, because of crowded conditions were limited to half-day sessions of work.

In using this material the teacher was able to take full advantage of the "laws of learning." She placed emphasis upon purposing and self-teaching, rather than upon the mechanical aspects of reading. Whenever a pupil found himself in a crucial situation he asked and received help from the teacher, but the real responsibility of his progress rested upon the child himself. He worked largely under his own direction, teaching himself the new words needed and learning to read in his own way according to the peculiar workings of his own mind.

For each group using the Practice Tests, a control group using another method, was selected. In each case this control group was equivalent to the experimental group in the following respects: type of school, ability of teacher (as nearly as it is possible to judge), number, age, intelligence, nationality, previous training, type of children, and the amount of time devoted to reading.

The Detroit Vocabulary Test was administered at the beginning of the experiment to all children in both groups but from the total number only two pupils recognized any of the words. The scores of these two were later discarded. It, therefore, seems safe to assume that all scores used in the tabulated results were those of children who started with zero ability in reading.

The teachers kept a record of the number of pages and books read by each individual throughout the semester. At the end of the time five standard reading tests were administered: Gray Oral Reading Test, Haggerty's Tests 1 and 2, Detroit Group Vocabulary Test, and Detroit Story Picture Test.

After tabulating the results from the total number of pupils to obtain some general data, the individual children in the two groups were matched according to age and intelligence, that is the scores earned by a child in the experimental group who was

six years and three months old and had an intelligence rating of A were matched with the scores of a child of same age and intelligence in the control group. All cases which could not be matched were discarded and the total matched cases of each group were combined to form the two exactly equivalent groups —one, the experimental, and the other, the control.

All charts shown were based upon data obtained from the results of these two exactly equivalent groups. Figure 5 shows the median scores resulting from the administration of the standard reading tests to the two matched groups.

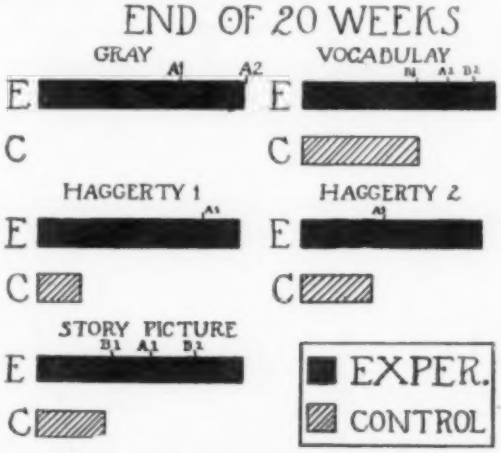


FIGURE 5. SHOWING THE MEDIAN SCORES ON THE STANDARD READING TESTS GIVEN TO TWO MATCHED GROUPS

In the Gray Oral Reading Test, the median score of the Experimental Group was 42.5 which is only 0.5 of a point below the standard for the grade II-A. The Control Group made no score on this test.

In the Group Vocabulary Test the Experimental Group earned a median score of 27.6. There is no standard for the grade II-A in this test, but pupils using the Practice Tests exceeded the II-B standard to such an extent that it seems safe to assume that they attained a II-A standard. The median score of the Control Group, for whose reading vocabulary the test had

previously been designed, was 16.8, which is slightly above the I-B standard and which indicates that they were a perfectly normal group.

In Haggerty's Test 1, the Experimental Group earned a median score of 4.8 which is above the I-A standard of 4. The median score of the control group on this test was 1.1.

In Haggerty's Test 2, the Experimental Group made a median score of 4.4 which is 2.4 points above the standard for the I-A grade. The Control Group made a median score of 1.8 in this test.

In the Story Picture Test the Experimental Group made a median score of 3.3 which is 2.5 above the standard for the II-B grade. There is no II-A standard for this test at the present time but again it is safe to assume that the Experimental Group attained or exceeded the II-A standard. The Control Group earned a median score of 1.2 which is the standard for the I-B grade.

The most significant fact to be gleaned from this chart is that in the most of the tests the I-B children using the Practice Tests made a median score at the end of five months which was equivalent to the II-A standard or the standard of children who had been reading for two years.

Figure 6 shows the median number of books read by each group in each level of intelligence. The numbers on the vertical scale denote the number of books read, while the horizontal scale denotes the different levels of intelligence. As is to be expected, in both cases, the median number of books read by the children of D and C intelligence was lower than that of the children of B and A intelligence.

In every instance, however, the median number of the books read by the Experimental Group was decidedly higher than that of the Control Group, the range in the latter group being from one book in the group of D intelligence to two and one-half books in the group of A intelligence, while in the Experimental Group the range of medians ran from four books read by children of D intelligence to eleven books read by children of A intelligence.

The most striking fact which the chart reveals is that the children in the lower quartile of the Experimental Group achieved

vastly more in the amount of material read, than did those in the upper quartile of the Control Group, the upper quartile of the pupils of D intelligence in the Control Group falling at 1.2 while the lower quartile of the pupils of the same intelligence in the Experimental Group fell at 3.5. A similar relation exists between these two quartiles throughout all the different levels of intelligence reaching the greatest divergence in the groups of A intelligence at which point the upper quartile of the Control Group falls at 2.7 books while the lower quartiles of the Experimental Group falls at 7.25 books.

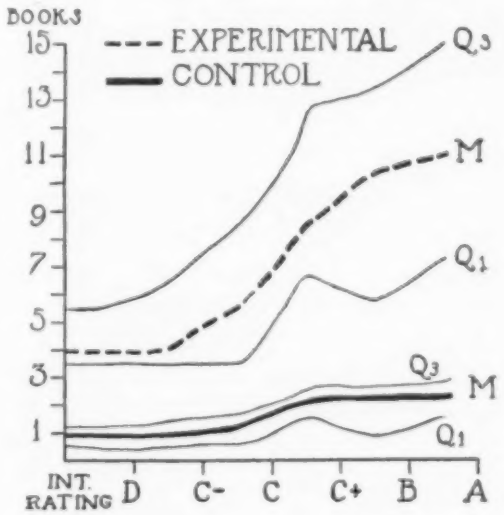


FIGURE 6. SHOWING THE MEDIAN NUMBER OF BOOKS READ BY EACH GROUP IN EACH LEVEL OF INTELLIGENCE

All first-grade children in Detroit are tested by the Detroit Group Intelligence Test and given one of the grades: A, B, C+, C, C—, D, or E. A is the grade given to children testing the highest, and E to those testing the lowest. The A's and B's are then placed together in one group called X's; the C+'s, C's, and C—'s are placed together in a group called Y's; and the D's and E's are placed together in a group called Z's. Thus the letter X is used to designate the group having the highest level

of intelligence; Y, the group of average intelligence; and Z, the children below average intelligence.

Figures 7, 8, and 9 show the distribution of the number and grade (primer, first or second readers) of books read by

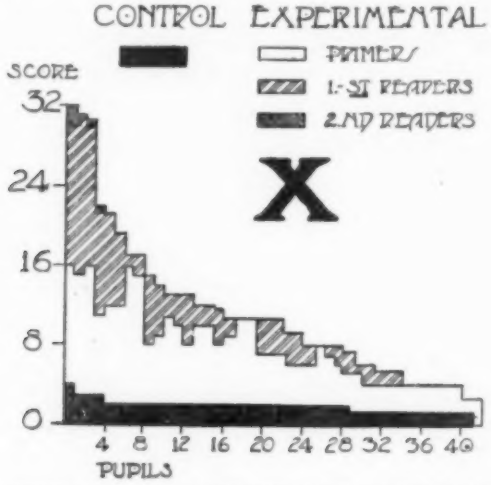


FIGURE 7. SHOWING THE NUMBER AND GRADE OF BOOKS READ BY MATCHED GROUPS OF X INTELLIGENCE

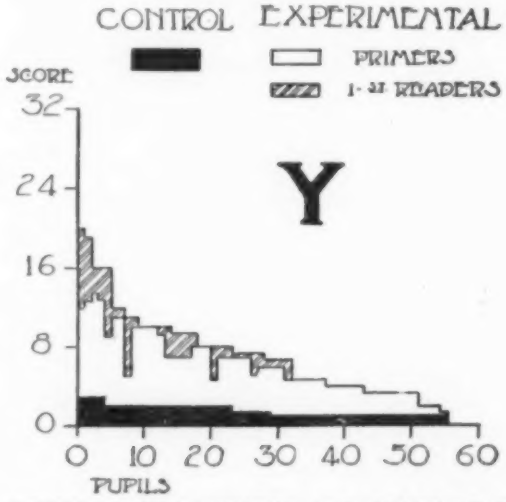


FIGURE 8. SHOWING THE NUMBER AND GRADE OF BOOKS READ BY MATCHED GROUPS OF Y INTELLIGENCE

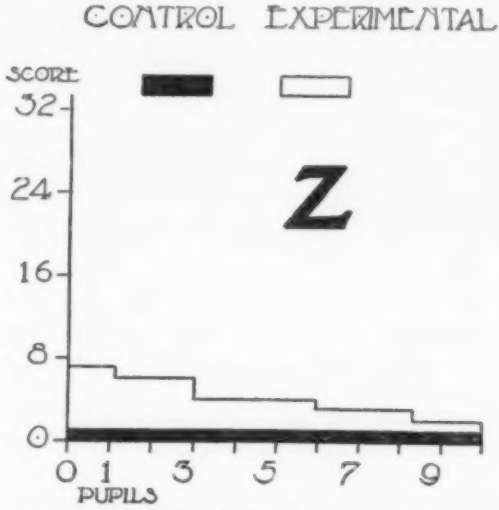


FIGURE 9. SHOWING THE NUMBER AND GRADE OF BOOKS READ BY MATCHED GROUPS OF Z INTELLIGENCE

matched groups of the X, Y, and Z levels of intelligence. In each of these figures the numbers on the vertical axis represent the number of books read, while those on the horizontal axis represent the number of matched pupils. The large black letter represents the intelligence rating of the group charted. The heavy line at the top of the chart shows the total number of books read by the pupils in the Experimental Group, while the line which marks the top of the blackened portion of the chart shows the total number of books read by the matched cases in the Control Group. The black and white surfaces represent primers read, the portion shaded with the oblique lines represents first readers read, and that shaded by the crossed lines represents second readers. Thus pupil Number 1 of X intelligence in the Experimental Group read 16 primers, 14 first readers, and 2 second readers; while pupil Number 1 in the Control Group, although of the same age and intelligence read but three primers during the semester. The range of the Experimental Group ran from 32 to 3 books, that of the Control from 4 to 1.

None of the children in the Y group read any second readers but many of those who used the Practice Tests read several first

readers. The range in the Experimental Group ran from twenty to two books read while that in the Control Group ran from two and one-half to three-fourths of a book read.

The children of Z intelligence read only primers in both cases but the number of primers read by the Experimental Group was much higher than that of the Control Group, ranging from 7 to 2 in the Experimental Group against one to three-fourths in the Control Group.

In no case did the teacher urge the child to read more books. Whatever was accomplished was done at the behest of the child's own "inner urge." In every instance, it is evident that in so far as achievement in reading is marked by the amount of material read, the children using the Practice Tests excelled by a wide margin.

Another significant feature to be noted is the wide variation in achievement which exists among the children of the same level of intelligence, when allowed to progress individually. To take the children of X intelligence, for instance; when obliged to follow the conventional classroom procedure, it will be seen that they all cover about the same amount of material during the term, but when allowed to progress according to their individual effort and ability, it will be noted that there are some "X's" who forge ahead reading as many as thirty or thirty-two books during the term while there are still others of this same level of intelligence who only read four or five books. This is not as high a number as that read by some of the children of Y or even Z intelligence. This evidence of the wide variation in the achievement of members of the same level of intelligence, and of the overlapping of the achievement of the members of different levels of intelligence, has been reproduced again and again in this and other experiments and constitutes one of the most convincing arguments for individual instruction.

Summary

Every one affected by this experiment seemed very well satisfied throughout the duration of its term. Parents were de-

lighted because their children were making such excellent progress; teachers were greatly pleased because they were convinced of the advantages of the method, and also because they were relieved of that everlasting "bugbear" of the first-grade teacher's life—"trying to find some seat work with which to keep the little ones busy"—and above all the children were happy because they were doing something which they wanted to do. Learning to read to them was certainly not "the scourge of infancy" as Rosseau once called it. To them it was a pleasure-giving and self-satisfying activity, the outgrowth of engaging in a pastime which had proved fascinating to little children ever since scissors and crayons were invented.

To sum up briefly the results of this experiment would indicate that with enough of this material to last throughout the semester, the Practice Tests in Reading might be expected to save at least two years in the child's reading life. There are two reasons, however, why it seems best not to use this data as final in drawing so important a conclusion: first, since this experimental material was only used in four schools, perhaps the number of cases is not large enough to be reliable, and second, since teachers in an experiment usually have special advantages in the way of contact with the supervisor. Since teachers often take a keener interest and put forth greater effort when trying out new material under experimental conditions, it is possible that they obtained better results than the teacher under ordinary circumstances would do. So for the time being it seems best to be conservative and say that *with the use of the Practice Tests in Reading the beginning first-grade child of average intelligence may be expected to read as well at the end of one semester as he has previously done at the end of one year.*

The essence of the entire experiment is embodied in a quotation from Mr. S. A. Courtis, Director of Instruction in Detroit Public Schools, under whose direction the Practice Tests in Reading were developed. They are but one expression of the general policy for which Mr. Courtis stands in Detroit and the writer had the benefit of his most excellent criticism and guid-

ance throughout the preparation of the material. Mr. Courtis says:

"Purpose is the key which unlocks every educational difficulty. So important is purpose that the best modern definition of education describes it as growth in power to form worthy purposes and to achieve them efficiently.

"This means that in teaching reading, the best results will be obtained, not by the teacher who works for results in reading, but by the teacher who sees in reading situations a new field in which to give her children experience in 'purposing' and 'achieving' and who develops 'power to purpose,' letting the reading results take care of themselves.

"The teacher, who in the spirit of service directs her efforts toward helping children carry out their own purposes in picture building and reading, will not only have the satisfaction of seeing her charges grow in personality, character, and intellectual power, but to these products, which are the essentials of all true teaching, will find added as a *by-product* that thing for which she did not work, namely understanding control of that most important of the school arts, the ability to read."

INTELLIGENCE LEVELS AMONG STATE NORMAL-SCHOOL GRADUATES

FREDERICK L. WHITNEY
State Teachers College, Moorhead, Minnesota

To a certain degree popular opinion has placed the elementary teachers in our public school systems on rather a low level of intelligence, and this opinion has seemed to be supported by the findings of a number of investigators who have reported since the World War.[1] The present inquiry deals with a sampling of that part of the teaching population which has graduated from state normal schools, and the results obtained point to the necessity of a revision of this opinion, at least as far as that part of the teaching population is concerned. This revision may be established as a proven fact by further study.

The graduates of the state normal schools constitute the elementary teaching corps in our larger school systems. This group, which comprises about one-fifth of the teachers of the country, has benefited from the definite attempt of the various states to give an adequate preparation for teaching in the public schools. If the state normal schools are admitting individuals of low mentality, there is grave danger that the objectives of public education may be defeated and that standards of the teaching profession may become debased. A number of tentative answers appear to the question, What levels of intelligence, as measured by accepted tests, are found among elementary teachers in our larger cities?

INTELLIGENCE LEVELS IN COLLEGE AND IN STATE NORMAL-SCHOOL GROUPS

The army tests were given to a group of 780 second-year students in eight state normal schools. The median score of 135

[1] Yoakum, C. S., and Yerkes, R. M., *Army mental tests.* New York: Henry Holt Company, 1920. 330 pp. Yerkes, R. M., *ed. Psychological examining in the United States Army,* (Memoirs of the National Academy of Sciences, Vol. XV), Washington, D. C.: Government Printing Office, 1921. 890 pp. Gambrill, Bessie L., "Some administrative uses of intelligence tests in state normal schools." *Twenty-first Yearbook of the National Society for the Study of Education,* 1922, pp. 223-43.

secured by this group may be compared with medians obtained by Van Wagenen from an examination of 887 freshmen at the University of Minnesota.[2] Using Form 6 of the Army Group Examination Alpha for eight of the university colleges he obtained the median scores given in Table I. The figures in parenthesis represent the medians stepped up to the amount of about eight points. The state normal-school students were advanced two years beyond the high school, while the college freshmen were but one. Detailed examination of levels of alpha intelligence found in groups at various stages of advancement from freshmen high school to graduate university seems to show increases of from five to ten points for each year.[3]

TABLE I. MEDIANS OBTAINED BY THE STUDENTS OF EIGHT COLLEGES AT THE UNIVERSITY OF MINNESOTA

Colleges	Medians
Chemistry	154.2 (162)
Arts men	133.3 (141)
Pharmacy	132.6 (140)
Engineering	131.6 (139)
Agriculture	130.0 (138)
Arts women	129.2 (137)
Home Economics	120.5 (128)
Dentistry	113.2 (121)

The central tendency of the eight college medians is practically the same as that found in the eight state normal schools.

A further confirmation of these facts is found in Table II in which are compared the medians for university and college students reported by a number of investigators[4] and the results

[2] Van Wagenen, M. J., "Some results and inferences derived from the use of the army tests at the University of Minnesota," *Journal of Applied Psychology*, 4:59-72, March, 1920.

[3] Van Wagenen, M. J., "Our schools as measured by the army tests." *Educational Administration and Supervision*, 5:163-76, April, 1919, *and* "Has the college student reached his mental maturity when he enters college?", *School and Society*, 9:663-66, May, 31, 1919. Yerkes, R. M. *ed. Op. cit.* pp. 544-46, 659-64.

[4] Yoakum, D. S., and Yerkes, R. M., *Op. cit.* Brown, J. C., *and* Maxwell G. E., "Intelligence of normal school students," *National School Digest*, 39:23, May-June, 1920. Reeder, R. C., "Application of army mental tests," *National School Digest*, 40:357-59, February, 1921.

from the eight state normal schools. Here the central tendencies for college groups seem to be from ten to fifteen points higher than those for state normal-school groups without accounting for any discrepancy which may be due to differences in stages of advancement. But, since the groups are such indefinite mixtures the influence of this factor cannot be determined accurately. The comparison of single figures for such distributions as these is not very profitable but the proportion of "A," and of "A" and "B" scores gives a better expression of the real differences.

TABLE II. A COMPARISON OF THE INTELLIGENCE SCORES AND CLASS RATINGS OBTAINED BY COLLEGE AND NORMAL SCHOOL STUDENTS

	Median	Percent of "A" scores	Percent of "A" and "B" scores
Universities	135	47	41
State Normal Schools	134	50	47
College Students	130	40	75
State Normal School Students	122	45	77
College Men	130	41	74
State Normal School Men	115	23	63
College Women	127	38	75
State Normal School Women	117	30	75

On the whole, no very great discrepancies appear except in the case of "A" scores gained by the two men's and the two women's groups. Here, the college men seem to have a distinct advantage.

College of Education Students from University Colleges and from State Normal Schools

A comparison of alpha levels in the case of the same group of 780 state normal-school second-year students and of 150 junior and senior students in a college of education gives the results shown in Table III.

TABLE III. ALPHA LEVELS OF STATE NORMAL-SCHOOL STUDENTS AND OF JUNIORS AND SENIORS IN COLLEGES OF EDUCATION

	College of Education	State Normal School
Median	152	135
Q_3	166	149
Q_1	138	114
Q	14.2	17.5
N	150	780

Thirty percent of the group of 150 are former state normal-school students. It will be recognized that this college group is probably the more highly selected, since they are advanced three and four years beyond the secondary school, while the group of 780 are but two years in advance. If an allowance of eight points be made here for selection or growth or some other determining factor, on the assumption that the college of education group were all seniors, the central tendencies would be identical. It will be noticed also that selection narrows the spread of the two distributions for Q, the quartile deviation, is smaller for the college of education group. The approximate identity of central tendencies is confirmed by a measurement of the intelligence of a group of 300 juniors in colleges of education by the Miller Mental Ability Test. The results of this measurement are given in Table IV according to the types of institutions which the students had previously attended. The figures for those who entered from other university colleges are very similar to those for students entering from state normal schools.

One hundred of these college of education students were checked by "honor points" earned during the two years of pro-

TABLE IV. INTELLIGENCE OF JUNIORS IN COLLEGES OF EDUCATION

	From University Colleges	From State Normal Schools
Median	97	96
Q_1	88	87
Q_3	104	101
N	223	67

fessional work in the university. The average number of honor points per credit hour was computed—an "A" being given three honor points; a "B," two; and a "C," one. It was found that former state normal-school students received more honor points in both academic and professional subjects. These facts are summarized in Table V.

TABLE V. HONOR POINTS EARNED BY JUNIORS IN COLLEGES OF EDUCATION

Students entering from:	Academic subjects: Honor points per credit hour	Education subjects: Honor points per credit hour
College of Science, Literature, and Arts.....	1.49	1.34
Minnesota State Normal Schools....	1.57	1.60

While this is not a measure of intelligence like Army Alpha, it may be that it takes account of larger amounts of influence coming from character traits and is, therefore, a better index for total mentality.

Occupation Level of Elementary Teaching

How does the intelligence of state normal-school graduates compare with intelligence levels in other professional and occupational groups? The Alpha scores obtained in the army have been tabulated by occupations,[5] and our state normal-school group may be placed here at its corresponding level.

First, however, it is interesting to note that on the basis of a distribution of scores according to the army literal notation, there is a close similarity between the standing of the 780 state normal-school graduates and the standing of army captains. Table VI brings these figures together. It also shows the distribution for 94,000 individuals of the white draft. The reader

[5] Yoakum, C. S., and Yerkes, R. M., *Op. cit.* p. 38.

TABLE VI. COMPARISON BETWEEN NORMAL-SCHOOL GRADUATES, ARMY CAPTAINS, AND THE WHITE DRAFT

Score	State Normal School	Army Captains	White Draft
A	47.4	53.4	4.1
B	38.2	29.0	8.0
C+	14.2	14.4	15.2
C	0.2	3.8	25.0
C−	0.0	0.4	23.8
D	0.0	0.0	17.0
D−	0.0	0.0	7.1
N	780	3,023	94,004

will note that only 0.2 percent of the normal-school graduates fell below C+.

If one uses the occupational ladder furnished by Yoakum and Yerkes, these literal groups appear as follows: "A" students, 47 percent of all, are on an intellectual level as high as or higher than the median engineer. "B" students, 38 percent of all, are on an intellectual level as high as or higher than the median physician, accountant, draftsman, dentist. "C" students, 14 percent of all, are on an intellectual level as high as or higher than the median bookkeeper, nurse, clerk, telegrapher, photographer, storekeeper, worker in concrete.

That is, about half of this group of prospective elementary teachers are found on a level of intelligence equivalent to that of army engineers, another forty percent may be compared with physicians and dentists, while the remainder, the smallest part, have literal ratings similar to those of bookkeepers, clerks, and skilled workmen.

The above facts, tentatively established, have many practical implications. Only two will be suggested here. From this study it would seem that the teacher material from which our state normal schools are making elementary teachers is very similar to that used by colleges of education in making secondary teachers. If so, this basis of essential equality may speed the happy day when matters of professional equipment and of compensation will be decided on grounds other than that of the age of the pupil groups taught.

Further, it may be that a skillful organization and management of our school systems should make certain that the capabilities of the entire corps of employees be capitalized in many realms of activity which have been, through habit or tradition, too specialized. In particular, there may be qualities of leadership among elementary teachers which have been neglected. Wider representation and fuller participation from this group will be safe, if it is finally determined that the intellectual levels tentatively established above are matters of real fact.

Addendum

A recent survey of all (10) Massachusetts state normal schools—Kirkpatrick, E. A., "Intelligence tests in Massachusetts normal schools," *School and Society*, 15:55-60, January 14, 1922—is, on the whole, confirmatory of the above statement. Kirkpatrick includes Thurstone's figures from Pennsylvania state normal schools and from thirty colleges, and says, "A larger percent of low-grade students is found in the freshman classes of colleges than in all classes in Massachusetts normal schools. Many of the low-grade college freshmen are doubtless eliminated before getting into higher classes of college. Colleges have and keep more brilliant students, yet between 80 and 90 percent of all students of colleges are of the same grade of intelligence as are found in the Massachusetts normal schools. This is quite contrary to the claims sometimes made, that bright students go to colleges and dull ones to normal schools." The writer used the Thorndike Intelligence Examination for high-school graduates with 101 seniors in an eastern state normal school and found a median score of 63. Thorndike found—unpublished data—the median for freshmen home-economics women in western colleges to be 59, for all freshmen in eastern liberal arts colleges 73, for all classes in southern liberal arts colleges 66, for all colleges of 'Type A' 76, and for all colleges of 'Type B,' 51.

SPELLING AGE COMPUTED FROM THE SCORE ON FIFTY PERCENT LISTS

WALTER E. MORGAN[1]
Principal, Webster School, Fresno, California

In a study of the accomplishment of school children, in which the quotient technique was used, it became necessary to determine the spelling ages of pupils of the elementary grades. Age norms were available for such product tests as the Woody-McCall Mixed Fundamentals, the Trabue-Kelley Completion Test Exercises, the Thorndike Visual Vocabulary Scales,[2] and the Thorndike-McCall Reading Scales. In spelling, however, there were no age norms. It was, therefore, necessary either to establish them or to adopt some technique which would be equally valid in determining spelling age.

Conversion of the grade norms for the fifty percent lists of the Buckingham Extension of the Ayres Spelling Scale into age norms, and the adaptation of Van Wagenen's[3] Table I so as to determine, from the pupils' scores on these lists, the point of fifty percent success and failure, would fulfill the need.

AGE NORMS FOR THE GRADES

After some search and communication some fifty thousand (50,456) pupils were located, distributed by ages as shown in Table I. The term 'year' signifies between nine and fifteen months, while 'half-year' means between three and nine months of age. Thus, in each case, the six-year group includes all pupils between five years nine months and six years three months of

[1] The author is indebted to Dr. Raymond H. Franzen for the suggestion of this technique.

[2] These norms, together with derivatives thereof, will appear shortly in two publications: Franzen, R. H., *The accomplishment ratio,* (Teachers College, Columbia University, Contributions to Education, No. 125.) New York: Teachers College, Columbia University; and *Measurement of reorganization,* edited by L. M. Terman, Chapter IV by R. H. Franzen, "The Conservation of Talent," published by World Book Company, Yonkers, New York.

[3] Van Wagenen, Marvin J., "Table for computing mean individual scores in educational scales," *Teachers College Record,* 21:432-40, November, 1920.

TABLE I. GRADE DISTRIBUTION AND MEAN AGES

Grade	Number of Pupils	Mean Age
I B	4687	6.24
I A	3516	7.04
II B	3689	7.47
II A	3289	8.12
III B	3641	8.80
III A	3214	9.33
IV B	3357	9.73
IV A	2873	10.18
V B	3467	10.69
V A	2870	11.41
VI B	3170	11.81
VI A	2648	12.36
VII B	2906	12.78
VII A	2594	13.01
VIII B	2405	13.66
VIII A	2130	14.12
Total	50456	

age; while the six and one-half year group includes those between six years three months and six years nine months of age. The cities in which the pupils are located, and the number in each, are: Napa, California, 1,096; Butte, Montana, 6,232; Rockford, Illinois, 6,928; Des Moines, Iowa, 15,434; and St. Paul, Minnesota, 20,766; total, 50,456.

When the average age for each yearly grade was obtained, it was found that the mean increment of age was 1.0357 years. This was then applied to grades beyond the eighth. The age norms for all grades are indicated in Table II.

TABLE II. AGE NORMS, GRADES I TO XII

Grades	1	2	3	4	5	6	7	8	9	10	11	12
Mean Ages...	6.64	7.80	9.07	9.96	11.05	12.09	12.90	13.89	14.93	15.96	17.00	18.03

Having established age norms for the grades, the next step was to evolve or adapt a technique which would result in spelling age.

Giving the Tests

Grades VI, VII, and VIII were selected for the purpose of establishing this technique. The method here given is, therefore, descriptive of the process used with these grades. It was extended, however, so as to render it usable with grades as low as the second.[4]

Ten words were selected at random from each of Lists R, T, V, X, and Z of the Buckingham-Ayres Spelling Scale. These are fifty percent lists for grades IV, V, VI, VII, and VIII. To the eight words of List AB were added one word from each of Lists AA and AC, giving a fifty percent list of ten words for grade IX. These sixty words were then dictated, each word being repeated in a sentence or definition. Each list of ten words was dictated separately, the pupils being instructed to place the proper literal designation at the head of each list.

When collected, the papers were rated by the examiner. All words actually misspelled or omitted were marked by a check, the number of such misspelled or omitted words being placed at the head of each list. Erasures or changes in spelling were not considered errors if the last spelling was obviously correct.

Computing Spelling Ages

In order to determine which would be the more economical administratively, two methods of scoring the tests and computing the spelling ages were used; that employed with the Trabue-Kelley Completion Test Exercises for determining individual performance levels,[5] and that proposed by Van Wagenen as a shorter way of accomplishing the same result. The latter method was finally adopted, since it eliminates a number of involved operations, which are time consuming and increase the probability of error.

---

[4] Dr. George C. Kyte, Principal of the University Elementary School of Berkeley, California, has employed this technique in grades II to V, inclusive, using lists L, O, R, T, V, and X. Results were very satisfactory in all but the low-second grade, where many pupils failed completely to spell any of the easiest list of words.

[5] Kelley, Truman Lee, "Individual testing with completion test exercises," *Teachers College Record*, 17:371-82, September, 1917.

Spelling age can be computed with considerable facility by employing an adaptation of Van Wagenen's Table I and using his Table II. Table I to which we refer is a table for calculating mean scores at the point where just fifty percent of the tasks are correctly done. It consists of scores, in terms of the difficulties of the tasks performed, which correspond to any number of errors, from one to ten, made at any scale step, from one to fifteen, on an educational scale. Since, however, this table presents scores from performance at a given scale step, it is necessary to make correction for errors on less difficult scale steps. This is done by subtracting from the score obtained one-tenth the sum of the errors made on less difficult scale steps. Van Wagenen's Table II is a table of further corrections to be made when two or more errors are made on the *least* difficult scale step.

Since his Table I is in terms of whole-number scale steps with equal intervals, and since the age values of the fifty percent spelling lists are non-integral, with unequal intervals, it was necessary to adopt the Van Wagenen table in order to arrive at the proper scores.

Table III reproduces the section of his table which was used and Table IV is the resulting adaptation. This was made first by substituting for the "scale steps" (i.e., for 7, 8, 9, etc.) the mean ages already shown in Table II—beginning, however, with the mean age (7.80) for the second grade. The next step in the

TABLE III. EXCERPT FROM VAN WAGENEN'S TABLE I

Errors	Scale Steps							
	7	8	9	10	11	12	13	14
1	9.5	10.5	11.5	12.5	13.5	14.5	15.5	16.5
2	8.6	9.6	10.6	11.6	12.6	13.6	14.6	15.6
3	8.2	9.2	10.2	11.2	12.2	13.2	14.2	15.2
4	7.8	8.8	9.8	10.8	11.8	12.8	13.8	14.8
5	7.5	8.5	9.5	10.5	11.5	12.5	13.5	14.5
6	7.2	8.2	9.2	10.2	11.2	12.2	13.2	14.2
7	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0
8	6.8	7.8	8.8	9.8	10.8	11.8	12.8	13.8
9	6.6	7.6	8.6	9.6	10.6	11.6	12.6	13.6
10	6.5	7.5	8.5	9.5	10.5	11.5	12.5	13.5

adaptation consisted in adding to the entries in each column of Table III the difference between the scale step of Table III and the corresponding mean age of Table IV.

TABLE IV. ADAPTATION OF TABLE III, FOR USE IN COMPUTING SPELLING AGE FROM SCORES ON THE FIFTY PERCENT LISTS OF THE BUCKINGHAM EXTENSION OF THE AYRES SPELLING SCALE

Errors	List L Grade II (Mean age, 7.80)	List O Grade III (Mean age, 9.07)	List R Grade IV (Mean age, 9.96)	List T Grade V (Mean age, 11.05)	List V Grade VI (Mean age, 12.09)	List X Grade VII (Mean age, 12.90)	List Z Grade VIII (Mean age, 13.89)	List ABC Grade IX (Mean age, 14.93)
1	10.30	11.57	12.46	13.55	14.59	15.40	16.39	17.43
2	9.40	10.67	11.56	12.65	13.69	14.50	15.49	16.53
3	9.00	10.27	11.16	12.25	13.29	14.10	15.09	16.13
4	8.60	9.87	10.76	11.85	12.89	13.70	14.69	15.73
5	8.30	9.57	10.46	11.55	12.59	13.40	14.39	15.43
6	8.00	9.27	10.18	11.25	12.29	13.10	14.09	15.13
7	7.80	9.07	9.96	11.05	12.09	12.90	13.89	14.93
8	7.60	8.87	9.76	10.85	11.89	12.70	13.69	14.73
9	7.40	8.67	9.56	10.65	11.69	12.50	13.49	14.53
10	7.30	8.57	9.46	10.55	11.59	12.40	13.39	14.43

We have already referred to Van Wagenen's Table II. As we have rearranged it, it is presented as Table V.

TABLE V. CORRECTIONS FOR ERRORS MADE ON THE LEAST DIFFICULT LIST OF WORDS

Errors.....	9	8	7	6	5	4	3	2
Add to amount to be subtracted	2.1	1.3	1.0	.7	.5	.3	.2	.1

By using Tables IV and V the process of obtaining spelling ages is rendered extremely simple. The entire process is reduced to four or five steps, as follows: (1) add the number of errors made on lists easier than the last (most difficult) list of words attempted; (2) divide this sum by ten; (3) in case two or more errors are made on the easiest list attempted, find in Table V the amount to be added to the quotient obtained in (2); (4) find in Table IV the interpolated Van

Wagenen value corresponding to the number of errors made on the last (most difficult) list of words attempted; (5) subtract from the interpolated V. W. value obtained in (4) either the quotient obtained in 2, or the sum obtained in 3. *The remainder is the spelling age.*

For the reader's benefit a job analysis of the spelling age technique is here given, followed by a number of examples illustrating the method used.

A Job Analysis for Computing Spelling Age

Instructions for Giving the Test

1. From the Buckingham Extension of the Ayres Spelling Scale select at random ten words from each of the columns L, O, R, T, V, X, Z, and to the eight words in column AB add one word from each of the columns AA and AC. If the test is for no grade higher than the sixth, columns L, O, R, T, V, and X are sufficient. When the Ayres List is used in this latter case, to the eight words in Column X add one word from each of the columns W and Y.

2. Distribute ordinary spelling paper and use pen and ink in all classes where the pupils are accustomed to writing with ink.

3. Have the pupils place at the top of the papers the following items: 1. *Name;* 2. *Date;* 3. *Grade;* 4. *Age in years and months* (to nearest 15 days).

4. Dictate the words in lists of ten each, the easiest list first, *being sure to have the pupils head each list with its proper literal designation.* Repeat each word in a sentence or in a definition, in order to obviate confusion of similar words.

If Grades III to VI are being tested, use the list of ten words from *columns L, O, R, T, V, and X only.* If Grades VI to VIII are being tested, *start with column R.* This will minimize considerably the giving and scoring of the tests and will be found thoroughly sufficient.

5. Collect papers immediately after completing the dictation.

Instructions for Scoring the Papers

1. The teacher should score all papers, using great care in marking and computing results. This will insure her greater help when she needs to use the final results to improve pupils' spelling.

2. Mark each word omitted or misspelled by a check.

3. If in the teacher's judgment the final spelling of the word was correct, consider it so. Do not consider changes or erasures as errors, if the final product is a correct spelling.

4. At the head of each list of ten words, write the total number of words misspelled or omitted from that list.

Directions for Computing Spelling Age

1. Add the total number of errors and omissions (see instruction 4 for scoring) made on all lists *except the last list attempted.*
2. Divide this total by 10 (point off one place to the left).
3. If two or more errors are made on the easiest list attempted, find in Table V the amount to be added to this quotient (found in step 2), and add accordingly.
4. Find in Table IV the numerical value corresponding to the number of errors made on the last (most difficult) list attempted.
5. Subtract from this value the value found in step 3 (or if less than two errors were made in the easiest list, subtract the quotient found in step 2).

Examples Illustrating the Computation of the Spelling Age

Pupil 1 made the following errors:

List	R	T	V	X	Z	ABC
Errors	0	0	0	1	5	8

Then, following through the steps given for computing Spelling Age:

1. The number of errors made on all lists except the last (ABC) is 6.
2. One-tenth of 6 is .6.
3. No errors were made in the easiest list attempted.
4. The number of errors made in the most difficult list (ABC) is 8. Finding 8 in the Errors column of Table II and following across the row to the column under List ABC, gives a value of 14.73.
5. Subtracting .6 (obtained in 2 above) from 14.73 (obtained in 4 above) gives 14.13, the spelling age.

Pupil 2 made the following errors:

List	R	T	V	X	Z	ABC
Errors	5	8	10	9	10	10

The following are the steps:

1. The number of errors made on all lists except the last is 42.
2. One-tenth of 42 is 4.2.
3. Five errors made on the easiest list attempted (R) gives a value of .5 to be added to 4.2, giving 4.7. (Table I).
4. The number of errors made on the most difficult list (ABC) is 10. Finding 10 in the Errors column of Table II and following across the row to the column under List ABC, gives a value of 14.43.
5. Subtracting 4.7 (obtained in 3 above) from 14.43 (obtained in 4 above) gives a Spelling Age of 9.73 years, or 117 months, figured to the nearest month.

Pupil 3 made the following errors:

List	R	T	V	X	Z	ABC
Errors	0	1	2	7	9	10

Then, using the same method:

1. The number of errors made on lists easier than ABC is 19.
2. One-tenth of 19 is 1.9.
3. No errors were made in List R.
4. In List ABC 10 errors were made. Using Table II this gives a value of 14.43.
5. Subtracting 1.9 from 14.43 gives a Spelling Age of 12.53 years or 150 months, figured to the nearest month.

The spelling-age technique as given in this article was evolved for use in grades six, seven, and eight. It may be used, however, in any grade for which there is a fifty-percent list in the Buckingham-Ayres Spelling Scale; namely, with grades II to VIII. In giving the lists to any grade or grades, it is advisable to give enough lists to test thoroughly the abilities of all the children. Words should be given which as nearly as possible will result in perfect scores on the easiest lists and zero scores on the most difficult lists. If the children are well-classified, three lists should suffice for any one grade; the fifty-percent list for the grade tested, and the fifty-percent list for the grades immediately above and below it.

A CONTROLLED EXPERIMENT TO DETERMINE THE EXTENT TO WHICH LATIN CAN FUNCTION IN THE SPELLING OF ENGLISH WORDS

WARREN W. COXE
Ohio State University

One of the problems which is being investigated by the Advisory Committee of the American Classical League in connection with the Classical Investigation is the influence of the study of Latin on the ability of pupils to spell English words. The investigation of this problem aims, first, to discover to what extent, if any, the study of Latin contributes to the pupil's ability in English spelling under present conditions of teaching Latin, and second, to determine under what conditions the greatest achievement can be obtained.

The investigation follows two lines: a general survey, participated in by about 125 schools from all parts of the country, in which the relative gains of Latin and of non-Latin pupils in spelling will be measured; and a controlled experiment, in which the conditions of transfer will be studied more in detail.

The controlled experiment is being carried on by the writer under the supervision of Dr. B. R. Buckingham, director of the Bureau of Educational Research of Ohio State University. Four schools in Rochester, New York, involving about 1400 pupils, and ten schools in Columbus, Ohio, involving about 1200 pupils, are cooperating.

The controlled experiment was begun in November, 1922, and will continue till May, 1923. It will use both Latin and non-Latin pupils from grades in which Latin instruction is begun. In some of the schools, Latin is begun in the seventh grade and in others in the ninth grade.

Five groups of classes are being used, each pursuing a different type of instruction. They are as follows:

Group I. Latin Control Group.—Classes in beginning Latin which will take the tests but otherwise pursue the ordinary course of study, making no effort to relate the Latin to English spelling.

Group II. Latin Experimental Group Alpha.—Classes in beginning Latin which will take the tests and follow a prescribed course of training. Each week of the training period, five English words will be taught. The teacher is to point out the relation of the Latin root to the English spelling, but is warned not to develop principles or rules.

Group III. Latin Experimental Group Beta.—Classes in beginning Latin which will take the tests and use the same material as that given to Group II, but will develop principles. In the instructions to the teachers, seven rules are listed, which she is to develop in class. The first rule is: Double consonants in Latin words are retained in the English derivatives, except at the end of compounds. Example: the "r" is doubled in *currere,* consequently it is doubled in occurrence, but it is not doubled in occur.

Group IV. English Control Group.—This group is composed of English (non-Latin) classes of the same grade as the beginning Latin classes with which they are to be compared. They will take the tests, but will do nothing in spelling except such work as is ordinarily included in the course.

Group V. English Experimental Group.—These classes are chosen in the same way that classes for Group IV are chosen. They will have definite instruction in spelling and will use the same words for practice material as are prescribed for the experimental Latin classes.

To measure the relative gains made by these five types of classes, two tests have been devised: the Buckingham-Coxe Spelling Test and an Ungraded Test. The first consists of fifty words of known difficulty, half of which are from Latin roots and the other half from non-Latin roots. Of the twenty-five from Latin roots, ten are from Latin words found in Henmon's Latin Vocabulary,[1] a list of Latin words common to thirteen first-year texts. The other fifteen are from Latin words found

[1] Henmon, V. A. C. "The measurement of ability in Latin." *Journal of Educational Psychology,* 13:515-38, November, 1917.

later in the high-school course, according to the Lodge[2] list. There will be three forms of this test, each comprising a different set of equally difficult words. The first form, Form A, was given at the beginning of the experiment, Form B was given in February, 1923, and Form C will be given in May, 1923.

The Ungraded Test is a supplement to the Buckingham-Coxe Spelling Test. It contains one hundred words, all of which are derived from Latin words found in the Henmon list. The difficulties of these words have not been determined, but those which were obviously easy to spell were not included. Half of the hundred words appear in the practice material which is used by the experimental classes, the other half are different. This test was given at the beginning of the experiment and will be repeated in May, 1923. There will probably be a practice effect due to having the same words repeated, but the gains due to this ought to be the same in all groups, and thus not destroy the possibility of comparisons.

Following are some of the questions which it is hoped can be answered from the results of the experiment:

1. To what extent is there an automatic transfer from the Latin to the spelling of English words? That is, when the relation of Latin to English spelling is not definitely pointed out, can we be sure that a transfer will take place?

2. When an effort is made to relate Latin to English spelling, what method of teaching is most effective?

3. When the same effort is made to teach spelling in non-Latin classes, how will the results compare with those obtained in the Latin classes?

4. Is increased ability to spell confined only to those words which are taught, or does it spread to other words?

5. Does the relating of Latin to English spelling result in improvement only in words of Latin derivation; or is the transfer more general, resulting in gains in non-Latin words?

[2] Lodge, Gonzalez. *The vocabulary of high-school Latin.* (Columbia Contributions to Education, Teachers College Series, No. 9) New York: Teachers College, Columbia University, 1907. 217 pp.

The first two questions relate primarily to the problem of the American Classical League's investigation. Questions 3, 4, and 5 are of vital interest and grow naturally out of the conditions set by the experiment. They are included in this study because of the Committee's intention to study related problems wherever the data available lend themselves easily to the treatment of these problems.

On the basis of the initial spelling test, pairs of pupils will be selected. That is, the gain of each pupil in Group I will be compared with the gain of a pupil in Group II who had the same initial score in the spelling test. Following are some of the groups which will be paired in this way.

1. The Control Latin Group will be paired with each of the Experimental Latin Groups to determine whether spelling ability is increased when Latin is related to English spelling.

2. The Experimental Latin Group Alpha will be paired with the Experimental Latin Group Beta to determine whether the use of principles or rules results in any increased spelling ability.

3. The Control Latin Group and the Control English Group will be paired to determine whether there is any automatic transfer due to the teaching of Latin in the ordinary way.

4. The Experimental English Group will be paired with each of the Experimental Latin Groups to determine whether as great gains in spelling can be made in an English class as in a Latin class.

5. Inasmuch as the teachers of the Experimental Latin Alpha classes are left more or less free to determine their own methods, it will probably be profitable to compare the efficiency of the different methods, by pairing pupils and determining the relative gains.

In addition to the determinations based on the pairings above mentioned, some additional studies of the data may be made. For example, intelligence ratings will be available for all pupils, and it will be possible to determine whether gains are related more closely to a general ability as measured by the intelligence test, or to a specific ability in spelling.

Editorials

"MORE THAN INFORMATION"

The school has forgotten that personality is superior to information. School men today are facing the spectacle of a great social agency whose chief efforts are directed upon the acquisition of information and skill. It is quite apparent that in this labor considerable progress and efficiency are in evidence. In spite of the claims of critics of public education the development of the technique of instruction has been remarkable during the last two decades. Information and skill are taught more efficiently today than at any time in the past.

If, however, we could choose or had to choose between information or skill on the one hand and personality and character on the other, it is certain that we should select the latter. To be courteous, honest, and unselfish is better than to be erudite. A nation with fundamental traits of character well developed is much more secure in its future and happier in its present activities than is one which possesses the widest range of scientific information. Ideals stabilize and direct civilization and whenever civilization declines and falls the fundamental cause lies in the disintegration of its ideals. Information is a servant of character and not a substitute.

Fortunately we do not need to choose between information and ideals. We may have both. Our problem today is not one of choosing an alternative; it is merely a question of emphasis. We can develop fundamental qualities of character and at the same time give valuable information. Indeed without information character cannot be strong. Ignorance has never been a good matrix within which to develop personality. The highest type of civilization is one in which fundamental traits of character are served by exact and scientific information.

The school is at fault in laying the emphasis upon information. Character is being developed but it is being done incidentally. The cause of this placing of emphasis lies in the current belief that ideals should be taught incidentally and not systematically. Since it is not necessary to analyze character traits, to scrutinize them, and to weigh them, when they are taught incidentally, we do not analyze them. Because we do not have to list the traits, we let them grow according to nature.

The result is that they are left to chance and accident. Nowhere within this nation of over a hundred million people can we find an authorative list of fundamental ideals. Individuals have prepared more or less thoroughgoing lists but the attention of the nation has not been turned upon the problem with sufficient continuity and persistence to analyze the task and so we are left in the dark.

It is an obvious fact that if subject matter is not brought to consciousness and placed in cold type in courses of study it will not receive adequate attention. When courses of study are filled with pages of material upon arithmetic, geography, and history and no pages are devoted to ideals, it is inevitable that the major attention of the teaching body will be placed upon material which they see in print. One of the laws of emphasis is emphasis by mass and where no space is given to an important subject it cannot possibly secure the attention which its importance demands.

The remedy lies in the recognition of those facts. The ideals to be taught must be determined with the same care as the content of arithmetic or spelling. The materials through which they are to be taught must be given a prominent place in courses of study and textbooks. These ideals and the contributory material must be apportioned according to grades with due regard to initial instruction and secondary repetition and supervision. Ideals must appear as topics in the course of study.

Once this has been done they can be taught either incidentally or formally depending upon the convictions of those who instruct. Some people believe that systematic instruction in morals is a mistake. Others believe that it is necessary. If it

is taught systematically it is clear that the material must be worked out in systematic form. But it must be remembered that if ideals are taught incidentally they must still be worked out systematically for the guidance of the teachers who teach personality by the incidental method. When it is taught formally it may be put into the hands of pupils. When it is taught incidentally the material must be put into the hands of teachers at least.

The development of personality lends itself to as intensive culture as does any biological phenomenon. Domestication and culture tend to improve quality. This is a common fact of biological life. Equally is it true of personality. By isolating the traits of personality and concentrating upon their development it is possible to produce a development which is amazingly more rapid and efficient than natural unstimulated development can be. Personality yields as freely to instruction as does information or skill.

It is time for the schools to take stock of the situation and direct their attention and skill upon the development of personality to the degree that character and ideals are superior in importance to information and skill.

W. W. C.

Reviews and Abstracts

PARKHURST, HELEN. *Education on the Dalton plan.* New York: E. P. Dutton Company. 1922, 278 pp.

At this period when the attention of educators is centered upon supervised study, the socialized recitation, the project method, and scientific measurement, this new volume comes as a distinct contribution.

The fact that the plan advocated has been successfully used both in this country and in England gives weight to its doctrines. The basic principle of the Dalton Plan provides opportunity in the secondary schools and upper elementary grades for pupils to initiate and organize their own procedure. While a definite course of study is suggested, the methods of following it break away from the traditional daily time schedule, the formal recitation, and the time-honored examination.

The entire year's work in each subject is broken up into carefully gauged monthly installments or projects, known as "contract jobs" which the pupils agree to undertake. These monthly assignments are subdivided into weekly tasks, and the pupil is permitted to take these up in any order he chooses.

The departmental plan is followed. The room for each subject is equipped with needed apparatus and an abundant supply of departmental reference books. The teacher in charge of the room is a specialist who teaches only one subject. The pupil may go to the room set apart for the subject in which he feels an interest, and work there as long as he wishes. Each pupil budgets his own time. To prevent him from working too continuously on his favorite subject he is not permitted to exceed its monthly prescription till he has completed the assignments in all other subjects. A pupil may do a week's work in one subject before proceeding to another. Interest in one subject will thereby, Miss Parkhurst thinks, spread to others less enjoyed. Before the pupil leaves the work room he plots upon a graph sheet the amount of work covered, so that he and the teacher can tell at a glance how much of the weekly assignment he has done.

The Dalton plan also affords opportunity for helpful cooperation. The pupils work in small groups helping each other and receiving such aid from the teacher as occasion demands. The pupil thus forms the kind of relationships in his school life that he will have later in business and professional experience. The teacher becomes a co-worker with the pupils. She gains such intimate knowledge of their achievements that formal examinations are unnecessary. Her function is to supervise and guide the groups in their study. The recitation period is abolished and lessons are given only as progress in the group may merit them.

Individual differences of the pupils are discovered through standardized mental tests, and the amount of work demanded from each is gauged by his mental ability. Each pupil is required to do the quantum and quality of work commensurate with his intellectual capacity. In a period of one month the brighter pupils often do the work outlined for six weeks, while the less gifted may do only three weeks' work.

One of the most helpful chapters in the book is the one devoted to illustrative assignments. The book is so crowded with excellent suggestions that no teacher or supervisor can study the volume and incorporate its sane principles with his educational practice without having his efforts in the schoolroom appreciably improved.

CHARLES FORDYCE

Department of Educational Measurements and Research,
University of Nebraska

BURTON, WILLIAM H. *Supervision and the improvement of teaching.* New York: D. Appleton & Company, 1922. 510 pp.

The introductory sentence of the Preface states that this book is "an attempt to discuss concretely the problems and procedures of the elementary-school supervisor." In realizing this aim, the book mentions hundreds of the problems that the supervisor will meet in his daily work. It, also, suggests effectual ways of handling some of the multitudinous, varied, vexing, and often times baffling problems which every true supervisor has recognized.

The author's point of view is that of the scientist. In developing his subject he uses experimentally determined evidence, when it is available. When such evidence is not available—and generally it is not—recourse is taken to expert opinion. The author well recognizes that many of his generalizations are not final, but he advances them tentatively with the hope that they will stimulate supervisors to study their problems in a thoroughgoing scientific way. Excellently annotated bibliographies after each chapter are calculated to aid supervisors in further study of their problems.

The book "gets down to brass tacks." It is concrete and practical. It evinces the first-hand acquaintance of the author with the supervisory situation and it is written so that it can be easily read and understood. It is a book that can be used—a book that admirably meets a professional need in a field that has all too long been uncultivated.

Since it is written primarily for those who have the duty of supervising, it will make its strongest appeal to superintendents, principals, and general and special supervisors—whether in training or actual service. But discussing, as it does the principles of general method and giving, as it does, an insight into the problems of supervision, it should make a strong appeal to the classroom teacher as well as to the supervisor. Indeed, the book is one that can be read with pleasure and profit by all members of the profession from the university president to the teacher in the elementary school.

WARD G. REEDER

Ohio State University

STRONG, E. J., JR. *Brief introductory psychology for teachers.* Baltimore, Md.: Warwick and York, 1922. 241 pp.

Whatever Doctor Strong has to say is of importance to a growing audience. Not everything that is psychologically acceptable is pertinent for teachers. Any author of a psychology for teachers, has to select but little of much available material. The selections here made are in sharp contrast to prevailing custom. The content of Strong's book overlaps the educational psychology textbooks bound to be written during the next decade under the impetus of the new dynamic psychology more than it does the texts of the past.

This book sets for its task the building up of four main concepts in the mind of the reader: (1) the understanding of actions, including thought, in terms of Thorndike's "situation-bond-response"; (2) a concept of learning as adding bonds; (3) individual differences; (4) elementary statistics (averages, deviations, correlations) for use in grading, classification, and general educational thinking.

The book attempts with far more than indifferent success to approach the ideal teaching situation of "pupil active—teacher passive." The illustrations from the very start are striking, often dramatic, uniformly "pat." The diction is congenial, friendly, and the opposite of highbrow in the offensive sense. With a few exceptions the reading difficulty of the content is not great. The book is divided into thirty lessons of reasonable length. In developing a subject, the first lesson is often an experiment to be written up, the next a discussion of principles involved in the preceding experiment.

Learning, the *pièce de resistance,* is the meatiest part of the book. The author's treatment here is clearly superior. He might have stressed a little more the importance of certain characteristics of successful practice or exercise. Nor is it clear why, with his unusual knowledge of the differences between success and failure in industry, he had so little to say about motivation, interest, and the effect of satisfaction and annoyance in learning.

The treatment of individual difference is clear, to the point, and connected with real problems of the schools. The treatment of grading, although enough pages are consumed, leaves the reader entirely uninformed upon many important aspects of grading—how to construct reliable and valid examinations for instance. The whole topic of marking is unfortunately discussed with the university problem as a background. The students of this work are presumably more interested in elementary and high-school affairs.

The treatment of statistics is not worthy of the author. His decision to introduce the beginning student to the statistical approach to educational problems is, in the reviewer's opinion, excellent but his handling of the subject spoils it. No mention of "probable error" and no place for "reliability" are found. This is a mistake because thinking guarded by notions of reliability is of prime usefulness to the group for whom this book is prepared. The development of skill in computing correlation is tried for but never justified. The author wants the reader, very properly, to be able to understand the more frequent statistical terms in professional discussions

and at the same time teaches average deviation and neglects standard deviation entirely. This is a mistake, for his readers will hear of standard deviation far more often than of average deviation. He gives the formula, $1 - \frac{6 \Sigma D^2}{n(n^2 - 1)}$ for the coefficient of correlation r, which it is not.

The treatment of educational and psychological tests are hardly more than press notices. The last 45 pages of the context are "poor relations" to the rest of the book. The evidential material supplied from experimental work could have been strengthened by referring to more recent reports. References to data published since the first edition of this book (1919) are conspicuously absent. With the exception of these last 45 pages, however, the book will prove a stimulating and satisfying treatise to the student of general psychology and the lay reader as well as a valuable text for the teacher.

F. B. KNIGHT

University of Iowa

SHARP, DALLAS LORE. *Education in a democracy.* Boston: Houghton Mifflin Company, 1922. 154 pp.

To many readers the authorship of this little book is a guarantee in advance of excellence in style, whatever they may happen to think of the views that are presented. The quality of its style is, in fact, the most outstanding feature of the book. For the most part the content is fairly familiar, yet is is said in a way that justifies abundantly the retelling of the tale. The author dwells on the fact that education on this continent was regarded from the outset, as a public function and a public obligation; that the public school is the greatest possible safeguard for democracy; that education should develop individuality as well as social adaptability; and that education must maintain close contact with life and nature, if it is to teach men to speak with authority and not as the Scribes and Pharisees. These various topics are developed in the four chapters which constitute the volume and which are presented under the titles, "The National School," "Education for Democracy," "Education for Individuality," and "Education for Authority."

As a plea for democracy and for a type of education that takes account of the deeper and finer things of life, the book contains much to commend itself. It combines intimate appreciation with magnificent expression. Perhaps it is ungenerous to find fault with so competent and whole-souled a spokesman. Yet the reader is left with a general impression that the author is, after all, in partial bondage to a formalistic and traditional conception of education. In his zeal for democracy, he advocates "an undivided, general course, broadly human, broadly democratic" (p. 55), without sufficient realization of the fact that vocational interests may be made to serve the ends of culture and of democracy. The fact that his conception of public education takes little cognizance of individual differences is acknowledged, at least indirectly, but it is claimed that "the school has a mass work to do, a national function to perform—to educate for democracy; the education for individuality must be given us everywhere, but not in any school" (p. 93). This work must be left primarily to the home. Most significant of all, perhaps, is the

author's failure to recognize the meaning of science in education and for democracy. He seems wholly oblivious to the fact upon which Robinson's "Mind in the Making" places so much emphasis, viz., that science is a conspicuous example of the process by which intelligence must be liberated from custom and tradition and pressed into service for the betterment of human life.

In a word, Mr. Sharp's doctrine of education is warmly and generously conceived, but it is cast too much in the mould of a literary tradition. It passes too lightly over the educational experiences that have accumulated since the days of Rousseau. The significance of these experiences is that we must take a more painstaking account of individual differences and of the meaning of science, in order to convert our aspirations for democracy into solid achievement.

B. H. Bode

Ohio State University

Young, Kimball, *Mental differences in certain immigrant groups.* (University of Oregon Publication, Vol. I, No. 11, July, 1922.) Eugene, Oregon: The University Press, 1922. 103 pp.

This study was carried on under the direction of Professors Cubberley and Terman at Stanford University. More than 500 children in San Jose, California, with additional children from smaller neighboring towns, making a total of approximately 800 cases, were studied. The racial groups investigated were those of (1) North European ancestry (called Americans), (2) Italian ancestry, (3) Portuguese ancestry, and (4) Spanish-Mexican ancestry. Only children above 12 and under 13 years of age were included.

The data obtained consisted of the results of (1) Army Alpha test; (2) Army Beta test; (3) teachers' estimates of intelligence (two separate judgments); (4) teachers' estimates of school work (two separate judgments); (5) school classification; and (6) occupations of parents. The data were all obtained and handled by the author or trained assistants under his direction.

The Italians, Portuguese, and Spanish-Mexican groups were found much alike in essential respects, and were classed together as Latins to be compared with the North European descendants classed as Americans. When thus classed, Mr. Young found that (1) in Army Alpha, only 7 percent of the Latins exceed the median of the Americans; (2) in the Army Beta, only 16 percent of the Latins exceed the median of the Americans; (3) in teachers' estimates of general intelligence, the Americans exceed the Latins by one-seventh of the range of the rating scale from "very inferior" to "very superior"; (4) in teachers' estimates of school work, the Americans exceed the Latins by one-seventh of the range of the rating scale; (5) in grade location, the Americans excel the Latins by nearly two half-year grades; and in occupational scale, there is a rough relation between economic class and the mean intelligence of the children. These are striking differences. Mr. Young's very careful handling of his data, his scholarly attitude throughout, and his references to other supporting studies carry confidence in the validity of conclusions.

The study sheds a bright light on the question of the part played by language difficulty in the differences among racial groups, repeatedly found in the intelligence scores. By correlations between the several sorts of data, Mr. Young shows very conclusively that the language factor is by no means so great as is commonly believed, and that the differences in scores is much more largely one of native intelligence. This constitutes a genuine contribution.

Two adverse criticisms may be made. The measures of central tendency should be found with the distribution tables, pages 12 to 15. In many of the tables, measures of variability should also be given. It requires the distribution table with its measures of central tendency and variability all together to tell the reader what he wants to know.

The above criticism is minor. The following is more important.

In the judgment of the reviewer, chapters IV and V, have no place in such a publication. They treat supposedly of the applications of the data to certain problems of school organization and method, and of certain theoretical implications concerning culture progress. In fact, these two chapters form the vehicle for the author's statement of his own philosophy of education, and his beliefs regarding eugenics and related fields, with but slight relation to the findings of his study. The reviewer happens to be in agreement with the author on the most essential tenets held, but reading the splendid study reported in chapters I, II, and III played little part in bringing about that agreement. Chapters IV and V should have wide reading—far wider than will be accorded them when printed following the technical treatment of data in the study.

As a matter of sound practice in publication, studies such as Mr. Young's should not be mixed with theoretical discussions.

F. J. KELLY

University of Kansas

News Items and Communications

This department will contain news items regarding research workers and their activities. It will also serve as a clearing house for more formal communications on similar topics, preferably of not more than five hundred words. These communications will be printed over the signatures of the authors. Address all correspondence concerning this department to Doctor E. J. Ashbaugh, Ohio State University, Columbus, Ohio.

At the San Jose State Teachers College the Annual Conference on Educational Research and Guidance will be held on April 20th and 21st, 1923. Doctor Buckingham of Ohio State University will be the principal speaker from outside of the state.

It is reported that the collection of money among the pupils of the New York City schools exceeds a million dollars a year. An effort is being made to regulate this matter.

The Méras-Roth-Wood French Scale comes in the form of a 14-page pamphlet containing general tests in comprehension and vocabulary followed by grammar tests for each of the following grades: VIII-A, VIII-B, IX-A, and IX-B. Tests in cultural information and pronunciation are also included. Copies of this test may be obtained from Doctor Ben Wood, Hamilton Hall, Columbia University.

The Detroit Freehand Lettering Scale, Experimental Edition, includes a small booklet containing specimens of lettering ranging in quality from 80 to 39; a pamphlet containing directions for giving, scoring, and tabulating; directions to students and blanks for the lettering Test A, Forms I and II. These may be obtained from the Board of Education, Detroit.

Miss Etta A. Dunn, head teacher of the Andrews School, Pittsburgh, Pennsylvania, writes that the upper classes of that school are being tested "to find out definitely whether the project method functions more completely than the traditional method of teaching." This is an interesting instance of an experimental attack on a problem about which much unsupported opinion is being expressed.

In response to a frequently repeated request for lists of books suitable for the library of the small high school, the Department of Public Instruction for Michigan, together with the Library Extension Service of the University of Michigan and the Librarian of the Michigan State Normal College have

prepared "A Library List for High Schools." This suggested list contains about a thousand titles arranged in subject groups.

Beginning with February 15, 1923, the Library Division of the Bureau of Educational Research at Ohio State University, as part of its state service, will issue a bi-weekly circular announcing the new publications which are received by the Bureau.

Educational measurements would not be true to their progressive character if they did not utilize the most up-to-date methods for their dissemination. We have already heard of two instances where addresses on educational measurements have been broadcasted. One of them was by Doctor J. Cayce Morrison who smote the atmosphere between Cuba and Baffin Bay on October 30, 1922. The station from which he talked was WHAZ, Troy, New York.

The school-building program for Winona, Minnesota, proposed by Mervin G. Neale and Sigurd B. Severson of the University of Minnesota was based upon a detailed investigation of the existing school plant and a careful study of the population trends for the past forty years with probable forecasts for the next twenty years. The Strayer and Engelhardt Score Card for City School Buildings was used to score the present school plant.

One of the New York City Survey committees expresses the opinion that experimental work is going forward among the New York schools which, if it were done in a smaller city, would attract nation-wide attention. The failure of the New York school people to attend national and state meetings is offered as one of the reasons why this lack of recognition exists. There is another reason. The New York people lack a real medium of publicity—one in which the nature of their investigations may be competently reported, and one which will be distributed widely.

The Schorling-Clark Practice Exercises for Accuracy and Speed in the Fundamentals in Arithmetic are timed practice tests arranged for pupils of the seventh and eighth grades. The authors, Raleigh Schorling and John R. Clark of Lincoln School, Teachers College, have included short practice exercises on the arithmetic processes from simple addition through percent. These are so arranged that pupils may practice for a short time each day, check their own work, keep their own record, and try to attain certain goals. These practice tests are so arranged that the work may be done individually with a minimum amount of supervision from the teacher.

Dr. Carlton W. Washburne, globe trotter and individual instructionist, lately stopped long enough while in Berlin to write a note saying that he expected to be home about the first of March. He has been talking up the JOURNAL with enthusiasm, as our letters from Europe indicate. According to Director Lipmann of the Institute for Applied Psychology at Berlin, Washburne says our Journal is "das wertvollste Organ der pädagogisch-

psychologischen Forschung in Amerika." If Washburne said so, we'll say he said a mouth full!

The important work which has been going forward at Los Angeles has eventuated in the issuance of tentative courses of study. These have appeared as separate monographs, and the JOURNAL has received twelve of them. Mr. Arleigh C. Griffin of the Department of Educational Research at Los Angeles indicates that the tentative edition is now exhausted and that an additional printing will not take place before summer. Dr. Bobbitt is again working with the Los Angeles people—or was early in February—and they expect to go much further in the work before the new series of monographs is printed.

Perhaps some of our readers have forgotten the N. E. A. Commission on the Coordination of Research Agencies which was appointed about two years ago with Superintendent Newlon of Denver as chairman. Early in its deliberations the Commission concluded that an important means of coordinating research agencies would be the use on the part of bureaus of research, librarians, and individual workers of a standard system of classifying educational writings. Accordingly, a sub-committee was created under the chairmanship of Professor Thomas H. Briggs for the purpose of working out a classification system. This sub-committee met at Columbus, Ohio, on February 18, 1923, to formulate a report for presentation at the Cleveland meeting.

Miss Elga M. Shearer, primary supervisor at Long Beach, California, has lately devised a course of study for use with the weak division of the lower first grade. Mr. Ernest P. Branson, director of research at Long Beach, in transmitting his course of study asks us to put him in touch with people who have formulated courses of study for any grades where ability groups have been organized. The editor of the JOURNAL OF EDUCATIONAL RESEARCH will be glad to receive copies of such courses of study from any school system which is adapting instruction to the different intelligence levels within any grade and will bring such courses to the attention of the readers of the JOURNAL.

The words of the Baltimore Survey (Vol. II, page 66) in connection with the Bureau of Research of the public schools of that city have significance for every city whether it is maintaining such a bureau or not. The Survey says: "The money expended . . . . in this department can be considered as one of the best investments which it is within the power of the Board of School Commissioners to make. In the last analysis this department provides the measures of success of all other departments, and hence its importance can not be over-estimated. The fact that such a department is assuming at present ever-increasing importance in the large majority of progressive school systems in the United States is one reason why Baltimore should provide for its development so as to secure the highest efficiency."

In a pamphlet entitled, "The Effect of Attendance upon School Achievement" Mr. Charles W. Odell of the Bureau of Educational Research, Uni-

versity of Illinois, concludes that "on the whole attendance appears to be a factor conditioning achievement but not so weighty a factor as many would have believed." Summing up the evidence from a large number of cases it appeared (1) that the percent of time which a pupil attends school has a rather definite effect upon his achievement age at the end of the period under consideration, (2) that it has practically no effect upon his actual increase in achievement age unless his attendance record is very poor in which case it lessens it for the same semester and increases it for the following one, and (3) that it is fairly closely related to his average school mark for the same semester and somewhat less so to that for the following one.

On Friday, February 9, at a meeting of the New York Society for the Experimental Study of Education an extended discussion was held on the use of intelligence tests in school grading. The school people of New York City have been making extensive use of these tests lately. Some of them have been doing so for a long time; others did not fall into line until it became evident that the administration, including the Board of Examiners, was sympathetic to the use of these instruments. At the meeting to which we are referring various emphases were placed upon the insufficiency of the intelligence test as a sole basis for grading. All, however, seemed to be in favor of the use of the tests provided they are supplemented by additional information. This is in line with the conclusions to which other school people are coming throughout the country.

An interesting survey of the New York City schools is being conducted from the inside. Some of the findings of the Committee on grading and course of study (Associate Superintendent Shimer, chairman) have recently been reported in *The New York Evening Mail.* The following important facts were established. Out of 480 schools reporting, 479 promoted pupils at irregular intervals and 95 percent of the pupils so promoted sustained themselves in the advanced grades. Regrading of classes was practiced in 222 schools, rapid advancement classes were organized in 152, opportunity classes in 179, and rapid advancement groups within classes in 177. Intelligence tests were used by 88 schools last year to select bright pupils; and standardized achievement tests were given by 52 others. In 218 schools teachers' judgments were used as the basis for promoting bright pupils. The committee declares that testing is an essential part of teaching.

From a study of the 1,910 errors made by 230 children on the Thorndike-McCall Reading Test, Form 4, Mr. W. J. Osburn, Director of Educational Measurements, Wisconsin State Department of Public Instruction finds that: (1) Forty-seven percent of the errors were due to a total failure to comprehend, indicated by the fact that the child copies irrelevant words and phrases, or "shreds of meaning" from the paragraph. (2) In 26 percent of errors the child's answer was distorted by a preconceived idea. (3) In 15 percent of the cases the errors were caused by failure to respond to all the question. These included cases where the question was correctly understood but only partly answered. (4) In 6 percent of the cases the answer was made up of

a combination of unrelated elements. (5) Six percent were due to miscellaneous errors.

Following this summary of the errors Mr. Osburn gives in his circular "Graded and Diagnostic Paragraphs for Use in Silent Reading," a series of paragraphs to be used in remedial treatment. At the end of each section he discusses the diagnostic value to be gained from a study of the types of errors which occur in children's answers.

Doctor John L. Stenquist, director of Educational Research at Baltimore, reports that he is working on a complete set of statistical records designed to show the status and progress of all branches of the work of the school system. In another connection he urges the necessity of classifying school children. He points out that in 1920-21 eleven thousand children failed of promotion in elementary schools. At a unit cost of $52, this amounts to a cost of over $500,000 for a single year. On this point Doctor Stenquist has this to say: "Scientific research has adequately shown that the most powerful factor in cutting down the percentage of failures is the reclassification of pupils on the basis of ability. Failure has been cut as much as 50 percent and over in various experiments. Assuming that through the Bureau we can reduce the number of failures even 10 percent, it is apparent that a saving far greater than the entire cost of the Bureau can be made in this activity alone. This takes no account of the incalculable saving made in human misery by reducing the number of pupils which we annually label as failure."

Bulletin, Number 48, of the Bureau of Educational Reference and Research, University of Michigan, reports the results of the Henmon Latin Vocabulary and Sentence Translation Test, Test II, given in May, 1922, to the pupils of Latin in the high schools of ten cities of the state. Certain of the findings are summarized as follows:

"Comparison of the median scores attained in the different cities in both vocabulary and sentence translation tests indicates that those cities that made high scores on the vocabulary test also made high scores on the translation test. The Bureau has not had time to figure the correlation between these two sets of scores, but it is reasonable to suggest that the ability to translate well carries with it the recognition of word meanings. There is some evidence that the opposite of this statement, that being good in vocabulary means being good in translation, is not so universally true. Most of the Michigan cities were up to standard in vocabulary, but not so many of them were up to standard in the translation test. If the two sets of the Henmon standards are equally reliable, then it must be concluded that mere knowledge of vocabulary is insufficient to cause the Michigan cities to attain the expected standards in translation. Exact translation demands knowledge of grammar with its emphasis on number, tense, case, mode, sentence structure, etc., and also the ability to express the meaning in clear-cut forceful English."

In *The Scottish Educational Journal* for December 15, 1922, Doctor Robert Rusk of St. Andrews University, Scotland, replies to the critics of the Fife Intelligence Tests. At the request of the Fife (County) Educational Authority, Doctor Rusk, acting as Honorary Consulting Psychologist, pre-

pared a set of group tests which were given to the pupils who came up for the secondary-school entrance examinations last year. The tests which appear in the Fife Intelligence Tests were chosen after considerable preliminary testing and their order was determined by their correlation with the Stanford Revision of the Binet scale. At the time of the examination, these pupils were given the usual entrance examinations and the Fife Intelligence Tests. In the opinion of those in authority, "the introduction of intelligence tests into the Fife County examinations has led to an increase in the number of gifted children admitted to the higher grade and secondary schools of the county. It is thought that the former examination emphasized too strongly the results of memorizing to the disadvantage of reasoning force and judgment. The inclusion of intelligence tests, on the other hand, has drawn into bolder relief the natural intelligence of the child, and a result of this can hardly be despised in the process of selecting young people for more advanced studies."

Miss Theresa E. Baller, chairman of the English Department, McKinley Junior High School, Los Angeles, California, reports the results of the Thorndike-McCall Reading Scale distributed according to intelligence quotients. The latter were obtained by the use of the Terman Group Test. The usual six classifications according to intelligence quotients are set up and for each of these classifications the number and percent of pupils receiving indicated reading quotients are entered. Pupils having intelligence quotients above 140 have reading quotients in no case less than 105; those whose intelligence quotients are between 120 and 140 do not do as well although they overlap extensively both beyond those in the next higher and those in the next lower intelligence group. Similar statements may be made with reference to pupils in the other intelligence classifications. In general, reading ability varies with intelligence. There is, however, such a marked degree of overlapping that the only two intelligence groups which are absolutely distinct from each other on the basis of reading quotients are those which have I.Q.'s above 140 and those which have I.Q.'s between 70 and 80.

Miss Baller indicates that she is retesting all pupils whose reading and intelligence quotients show marked discrepancies and that she expects to make a special study of the cases which continue to show such discrepancy after the retesting.

Last spring an interesting contest took place in Cloud County, Kansas. Although this was nearly a year ago, the features of this contest have present interest. Director Dean A. Worcester, of the Bureau of Educational Measurements and Standards at Emporia, had charge of the project. It seems that in Cloud County, as in so many other counties, contests of athletic, musical, and literary ability had been held annually. Last spring the competition was extended to include academic subjects. The plan was to test the children in the morning and to have the papers graded so that awards might be made on the evening of the same day. Students entered the contest as they would

enter an athletic contest. Medals and cups were given for first, second, and third places.

The subjects in which testing was carried out were English composition, United States history, civics, first- and second-year Latin, algebra, geometry, manual training, home economics, and physics. Standardized tests were not used because many of the children were already familiar with some of them. The bureau of which Mr. Worcester is director undertook to provide the examination questions. Some of them were of the true-false type, others of the multiple-choice type, and still other of the completion type. From the nature of the case the examinations had to be easily, rapidly, and objectively scored.

The project was very successful indeed and attracted a good deal of interest throughout the section of the country in the vicinity of Cloud County. Mr. Worcester writes that he considers it "one of the most valuable pieces of work that we have attempted this year."

The experiment which is being carried on at Los Angeles, California, and through which children who are behind in their work are brought up to grade in the so-called "adjustment room" is attracting attention rather widely. The JOURNAL has already published a brief outline on the subject, but a much fuller treatment of it appears in the form of a handbook published by the Department of Psychology and Educational Research (Doctor A. H. Sutherland, Director). This handbook is designed for the use of teachers of adjustment rooms. A reading of it and an examination of the adjustment progress card impresses one exceedingly favorably. In some of its aspects we are reminded of the individualized instruction which Superintendent Washburne of Winnetka, Illinois, has been carrying forward. There is, however, a great deal that is original in the plan as it is being worked out at Los Angeles. The brief article by Doctor Sutherland in the September 1920 issue of the *Elementary School Journal* relates particularly to the reading work in the adjustment rooms. Doctor Sutherland also indicates in this article that an experiment is under way in the application of the special materials in regular classes. Half the children in grades II to V of a certain school are using the materials; while the other half follows the old plan. We shall be glad to read Doctor Sutherland's report of the experiment when his data justify him in publishing it.

During the Christmas vacation Dr. B. T. Baldwin, director of the Iowa Child Welfare Research Station, organized a comprehensive investigation for the public school system of Cleveland, Ohio. The progress of children is to be based on their physiological, mental, pedagogical, and social ages.

Typical schools in each of the five city school districts have been selected and the work will be carried out by the superintendents, the five supervisors, and the principals of the schools with the direct cooperation of the departments of instruction, educational research, psychological clinic, health, physical training, and nutrition. The program will include two semesters' work of

classes from kindergarten through the junior high school. No effort will be spared by the school authorities to make the experiment thorough and practical.

Mr. W. A. Justice, director of research at Virginia, Minnesota, has lately issued a report of the work done in his department last year together with an announcement of his plans for the current year. As to the work done last year, forty-five items are briefly described. These include not only testing activities but such other matters as pupil distribution, school census, practice materials, teaching load in the high school, salaries and cost of living, non-promotion and retardation, enrollment cards, under-weight pupils, and others literally "too numerous to mention." One is impressed by the unusual range of the activities of the Research Department at Virginia.

Mr. Justice writes that the introduction of a filing device involving the use of individual folders for pupils is probably the biggest thing his department has done. With particular reference to this filing system he says: "All records of achievement and intelligence tests—group or individual—are placed in the respective folders. In a year or two we shall have gathered in this file a fine body of information concerning each child, to be used for industrial-vocational guidance." Those who are interested in a record blank on which to carry the results of testing will wish to see the form used by Mr. Justice. The attendance and scholarship record of a pupil for a year is exhibited on one side of a 8½ x 11 sheet for loose-leaf binding. On the other side is the pupil's record according to standardized tests and scales. Since the record is intended to cover but one year it is easy to provide for every entry likely to be desired. Presumably at the end of the year the form is filed in the pupil's folder. Thus over a period of time the data on standardized tests for each child will be completely shown. The plan, however, fails to provide, as some have tried to provide, for a continuous test record which shall cover the school life of the child and which shall be exhibited on a single form.

The growth of teacher-training institutions has been especially observable in recent years. This growth is the resultant not only of the large increase in the number of high-school graduates—i.e., in the number of students prepared to enter the teacher-training institutions—but also of the very great number of teachers in service who are availing themselves of the additional training which these institutions afford.

No more striking evidence of the latter source of growth has come to our attention than that which is reported in *The School Magazine* (Buffalo) for December, 1922. According to an editorial in that issue, it is stated that the educational department neither urges nor enforces the use on the part of teachers in service of the opportunities for additional training. It is further stated that increments of salary have not been conditioned upon the taking of extension courses. Under these circumstances the record of extension work done by the Buffalo teachers is indeed remarkable. During the past four years they have earned credit for 16,815 semester hours. This does not include the work done by 130 teachers who during this time secured degrees, nor does it include attendance upon courses for which college credit

was received. When these two types of work are included, a grand total of 33,200 semester hours is reached. This is the equivalent of more than 1,100 solid years of college work. It amounts to 10 semester hours for each teacher in the Buffalo system, or a little more than an hour per week for each teacher throughout the entire four years.

If this amount of work had been accompanied by the usual costs of education, it would have represented an expenditure of more than two million dollars. As a matter of fact, it cost the teachers almost nothing in money. At the time the editorial to which we are referring was written, 82 percent of the teachers in the Buffalo system were reported as having taken some kind of work for professional improvement during the four-year period under consideration.

Superintendent C. C. Underwood of Painesville, Ohio, has combined the results secured from the Illinois Examination with the success of pupils in their school work to form ability groups. So far as this procedure relates to the grades below and above the seventh and eighth, it does not appear that Superintendent Underwood's arrangements materially differ from those which have been reported by other people.

In the seventh and eighth grades, however, he is using an unusual plan. In each section—i.e., in VII-B, VII-A, VIII-B, and VIII-A—pupils have been grouped according to the amount of work they are capable of doing, without regard to intelligence scores. Those who are doing exceptionally good work receive more exacting assignments, their classes are large, and the amount of individual instruction is relatively small. Those who are doing average work receive easier assignments although, in general, they cover the same ground. The sections are smaller, and more individual help from the teacher is available.

The most distinctive feature of Superintendent Underwood's plan is the method by which he handles the remainder of the pupils in the four half-yearly grades from VII-B to VIII-A, inclusive. All of these pupils without respect to grade are put in one group called Class A-1. Most of these pupils are over age and mentally below normal. Their school work is unsatisfactory. They have an entirely different course of study from that of the rest of the pupils. It is taught, in the main, by the project method; and a maximum degree of individual assistance is given by the teacher. There are no recitations; and all periods are really supervised study periods. These pupils are promoted without regard to the kind of work they do, provided the teacher feels that they have done their best. In other words, a VII-B pupil is promoted to VII-A at the end of the semester if he has made reasonable effort. The same is true of pupils in VII-A, VIII-B, and VIII-A. Pupils promoted from VIII-A, however, go to the ninth grade as special pupils and are not regarded as prepared for the regular high-school work. Their curriculum consists largely of manual arts and allied subjects.

The distinctive feature of this grouping is the organization into one class of the subnormal pupils of the four half-yearly grades and their promotion on the basis of effort.

At Seattle, Washington, there has been a movement directed toward the establishment of an educational survey. This year the Board of Directors established a Department of Research for the public schools, and Professor Fred C. Ayer of the University of Washington was appointed director. The Board in conference with the Superintendent laid down a program of topics for special study. One of these had to do with the organization of the school system.

Accordingly, Professor Ayer sent a circular letter to certain educational people asking them to judge each of four types of organization by using ten criteria. The four types of organization were named and defined as follows (the first three already exist in Seattle and the fourth is being considered):

I. *Academic.*—The conventional organization in which approximately forty pupils are taught by one teacher in one room throughout the school day. The pupils are divided into two divisions which usually alternate in recitation and study or special occupation. The teacher is assisted by an occasional visit from special supervisors.

II. *Industrial.*—Essentially the same as the Academic in the first six grades. In the seventh and eighth grades from sixty to seventy-five minutes daily are given to girls' home economics and boys' shop work taught by special teachers in a special room. Time for this industrial work is taken from the academic subjects.

III. *Duplicate.*—The school work of each classroom group of children from grades IV to VIII is organized into two equal parts, (a) academic and (b) special. The academic work (reading, arithmetic, etc.) is taught by one teacher in the home room while the special work (music, manual training, penmanship, etc.) is taught by special teachers in especially equipped rooms. Two groups of children shift between the home rooms and the special rooms each quarter of the day. The school day is lengthened and more time and emphasis are given to special subjects and activities than in the academic schools. Teachers ordinarily have but a single division of pupils at a time. The division, however, is usually larger than the recitation division of the Academic grouping and may be several times larger with certain of the special subjects.

IV. *Junior High School.*—A plan of reorganization which groups grades VII, VIII, and IX (high school) in one building. The work is semi-departmentalized upon the duplicate plan, promotion is by subject, the curriculum is broadened, courses are differentiated more widely and more attention is given individual variations, guidance, and social organization.

The recipients of Professor Ayer's letter were asked to judge these four types of work with reference to the ten following points: attendance, progress, achievement, curriculum, instruction, supervision, administration, guidance, socialization, and low cost. These ten items were carefully defined in the questionnaire.

Professor Ayer says that the general topic of organization is particularly pressing at Seattle owing to the impending building program. We hope that he will make his findings generally accessible not only because of the wide interest in them but also because building programs are contemplated in many other cities. On the basis of his returns, Professor Ayer indicates his intention of making a quantitative study of the ten factors (attendance, progress, etc.) as they may be found in the three types of organization now in operation at Seattle.

We feel that this study is important and we hope that in due time a report of it will be made available.

There is a great deal of present interest in the building principal. One form which this interest is taking is that of inquiry into what he or she is doing. We wonder what our readers will think of the following "report of activities for one day by a grade principal."

About 8:20. Arrived at the building. Marked a set of papers handed in Monday evening. Received money from children for tickets and costumes. Checked this on list. Sold a book. Sold some paints. Made change for both children. Talked with Miss J. about L. J. who is not doing good work this semester. Helped Mrs. B., a substitute teacher, plan work for the day. Looked up in our Classification List the set of lantern slides needed for a lesson soon. Made out the list of supplies I wished the janitor to get from the office. Heard a complaint from E. H. about R. B. hurting her with a snowball. Investigated—heard both sides. Returned papers I marked and talked with pupils about them. Talked with Mrs. B. and Miss J. about three children who were absent. Signed several bank cards and took them to the rooms.

9:30-10:00. Gave music lesson to Room 11 and Room 12 combined. Collected some Christmas music we had used, marked it, and filed it away. Sorted out some books I had used for reference on Indian costumes, and returned to teachers. Returned some reference books to three rooms. Arranged some textbooks in office preparatory to counting for book report.

10:30-10:45. Watched boys go to the basement and out for recess. Boys reported something wrong with merry-go-round; had janitor fix it. Looked around for and asked two teachers about an old Aldine Chart that Miss J. asked for. Reported to Miss J. Made out an alphabetical list of names in Room 11 to be used in making index for permanent record. Mr. H., manual-training teacher, called about some plans for next week's work. Conferred with Miss B., and reported to Mr. H. Went to three rooms and consulted with teachers about the spelling of some pupils' names over which there was a disagreement in the lists sent in for permanent record index. Consulted with Miss G. about some new material for our "play store." Looked through some books to find a certain story about Booker T. Washington. A patron of the school had asked me to find this story for her to use in a club meeting. Examined M. R.'s arithmetic paper and recorded her grade. I do this every day because her mother was not satisfied with her marks last month. Made out outline for reading lesson and gave to Miss J. for her to put on board. Looked through records of a teacher who is absent to find data for a card to fill out for a boy who is leaving town. (Report card for some reason had not been filled out.)

12 noon. Went home. Met Mrs. H. and talked with her about D.'s work in school.

12:45. Returned to building. Arranged lists of names sent in by teachers for index. Talked with Mrs. J. (a substitute) about some pupils who are not doing good work, and also about an absentee. Heard a complaint from some girls who were hit by snowballs. Investigated and reported to boy's teacher. Examined a little girl in Room 1 who looked ill and had fever. Sent her home. Visited Rooms 1, 2, 3, and 4 to find a little girl whose mittens had been taken. Found the child who did it and sent her for them. Looked through Miss M.'s desk to find her bank cards and helped Mrs. B. (a substitute) fix the cards for the day.

1:40-2:10. Gave a fifth-grade reading lesson in basement. Had an interview with Mr. B. about some bills he wanted given to Miss C.

2:15. Returned to office and found Miss A. from Y. W. C. A. who came to talk with me about Girls' Reserve for next semester. Visited Rooms 7, 8, 9, 10, and 11, to find how many girls are twelve years old or more. Sent a boy to office with report.

2:30-2:45. Watched boys from six upper rooms go to and return from basement. Took a little girl's temperature, talked with her, and put her to bed on the couch in the office when I found her mother was not at home. Talked with a boy who is doing poor work in school. Wrote his mother a letter explaining what I think is the cause and asking parents' cooperation. (Mother is deaf so I can't talk to her over the phone.) Signed some bank cards and took them to rooms. Went to Rooms 9, 10, and 11 to announce a meeting of Girls' Reserve after school.

3:30. Counted bank money—$26.03. Signed more bank cards and fixed up my record. Took money to Janitor to conceal some place in building as I didn't get it counted in time to send to bank before three. Talked with the girls who stayed to attend meeting of Girls' Reserve. These were prospective members and all had questions to ask. Mrs. W. came to conduct the meeting. Talked with her about plans for next semester. Made plans with Mrs. B. about her sewing lesson for tomorrow.

4:25. Left the building.

4:30-5:50. Attended Teachers' Club at A school to hear lecture on Passion Play.

During the day I answered the phone a great many times but can't remember when nor what the calls were.

Arithmetic Skills of Prospective Teachers

The academic scholarship of teachers in training has often been the subject of animated discussion. Teachers in colleges of education and normal schools are wont to say that their students know little of history, geography, and arithmetic, and that their skills in reading, spelling, English composition, and arithmetic are far below desirable levels.

On February 9, 1923, the writer gave the Courtis arithmetic tests, Series B, to 104 freshman women,—students in his course, "Teaching Arithmetic in the First Six Grades." Their median rate scores were: addition, 10.3; subtraction, 13.2; multiplication, 11.4; division, 10.8. The eighth-grade rate standards, set by Courtis, are: addition, 12; subtraction, 13; multiplication, 11; division, 11. It will be seen that the median scores of these students are practically coincident with the eighth-grade standards, except in addition. In this operation they rank as sixth graders.

The median accuracy scores, expressed in percents, were: addition, 65; subtraction, 83; multiplication, 72; division, 85. The eighth-grade standards, as reported by Courtis from testing done in 1916, are: addition, 76; subtraction, 87; multiplication, 81; division, 91. It is evident that these 104 college students are less accurate than eighth-grade pupils in each operation. In fact the accuracy medians are practically identical with those attained by fourth-grade pupils in addition, fifth-grade pupils in subtraction, and sixth-grade pupils in division, while in multiplication the median is just midway between the figures reported for pupils of the fourth and fifth grades.

The writer has frequently asked school men what they thought should be the minimum degree of skill required of prospective teachers in the four fundamental operations with integers. No one has been found but would

require eighth-grade median performance or better in each operation. Suppose that we apply this standard to our 104 freshmen.

Of the 104, the addition rate standard is attained by 28.
Of the 28, the addition accuracy standard is attained by 9.
Of the 9, the subtraction rate standard is attained by 8.
Of the 8, the subtraction accuracy standard is attained by 3.
Of the 3, the multiplication rate standard is attained by 3.
Of the 3, the multiplication accuracy standard is attained by 1.

This one reaches the division standards in both rate and accuracy. The imposition of the eighth-grade median standards would eliminate 103 of the 104 recruits.

We can tell the story in another way. The standard for rate was attained by 28 in addition, 54 in subtraction, 57 in multiplication, and 48 in division. The standard for accuracy was met by 27 in addition, 38 in subtraction, 26 in multiplication, and 34 in division. The number of students attaining both rate and accuracy standards was 9 in addition, 22 in subtraction, 23 in multiplication, and 25 in division.

R. L. MORTON

Ohio University

On Doctor Bagley's Rejoinder

While one hesitates to intervene in the gladiatorial contest of East and West over the I.Q. and its significance, or to do anything that might hasten the dialectical display that it has occasioned, the following comments on Doctor Bagley's rejoinder in your December issue may not come amiss and may help to clarify the issues.

Doctor Bagley assumes that readers will infer that because all were capable of utilizing writing once it was invented, therefore all were capable of inventing it, that the ability to apply a discovery is equivalent to the ability to make the discovery. The users of motor cars and of wireless are not so presumptuous. This instance of writing is illustrative of Doctor Bagley's general principle. Even in democratic communities the common man does not originate policies; he merely accepts or rejects them. And the better the policy which he has no part in originating, the better will be the rule of the common man. The world's present difficulties may not be due to the rule of the common man but to the fact that some countries do not possess individuals with high enough mentalities to originate satisfactory policies. The education of America is likewise not governed solely by the collective rule of the common man, but is largely inspired by the educational aristocrats of Teachers College, Columbia University. If Doctor Bagley really believes in his own doctrine, he ought to give up trying to influence American educational opinion and submit to the collective rule of the common school man.

Doctor Bagley also attributes to the mental tester a doctrine of heredity which the latter does not accept. Doctor Bagley implies that the mental tester believes that the I.Q.'s of Adam and Eve have determined the I.Q. of each

succeeding generation and that this has throughout the ages remained constant. But the Terman I.Q. is an I.Q. of individuals, not an I.Q. of the collective common man. What the mental tester affirms is that the I.Q. of an individual is innate. He does not seek to explain its origin or to fix responsibility for its possession. He would not be surprised to find that even when both parents have an I.Q. of 100, the I.Q.'s of their children were distributed thus: 80, 90, 100, 110, 120. All he would say is that the child with the 80 I.Q. would not go as far as the child with the 120 I.Q. The mental tester leaves the explanation of the differences in the I.Q.'s to speculative philosophers like Doctor Bagley. Doctor Bagley assumes that the horizontal growth or reach of the 80 I.Q. individual might compensate for the vertical reach of the 120 I.Q. individual; but the assumption that the horizontal reach might correlate directly with the vertical reach is more likely to be valid, as Professor Thorndike will explain to Doctor Bagley. The relative "low visibility" of the difference between the common man and the genius which Doctor Bagley regards as a "crucial instance," may be due to the fact that the education provided for the genius has not been adapted to him but rather to the common man. In cases like J. S. Mill, where the education was specially devised for his higher capacity, the difference was more clearly manifest. The mental tester has not investigated whether the general average of I.Q.'s from generation to generation remains constant, and he is consequently not likely to make dogmatic statements about it. Any view of this question attributed to him by Doctor Bagley is unjustified.

The correlation between the army tests and educational facilities cited by Doctor Bagley may be due to the fact that as a country gets the government it desires, so a state may get the educational system which it has the capacity to appreciate and benefit by, more especially when it is subject to the collective rule of the common man. What Doctor Bagley ignores throughout is that democracy is not an existing entity, but constituted out of individuals with different capacities, and he seems to forget that, as T. H. Green has said, "There can be nothing in a nation however exalted its mission, or in a society however perfectly organized, which is not in the persons composing the nation or society . . . . The spiritual progress of mankind is an unmeaning phrase unless it means a progress of personal character and to personal character . . . . It is simply unintelligible unless understood to be in the direction of more perfect forms of personal life."[1]

ROBERT R. RUSK

St. Andrews University, St. Andrews, Scotland

Educational Research in Music

Educational research in music is a pioneer field. For example, the Hillbrand scale for measuring ability in sight-singing contains the only standardized educational tests in the field of vocal music. There are indications, however, that in the next few years many valuable educational researches

[1] Prolegomena to Ethics, §§ 184, 185.

in music will be produced. At present there are only a few workers in the field. The only tests now available are the following:

Music Intelligence Test by Glenn Gildersleeve of Washington Junior High School, Rochester, New York. The work was done at Columbia University.

Beach Standardized Music Tests by Frank A. Beach of the Kansas State Normal School, Emporia, Kansas.

Detroit Practice Tests devised under the direction of S. A. Courtis and Thomas H. Chilvers, supervisor of music.

Standardization Tests by Charles A. Fullerton of Iowa State Teachers College, Cedar Falls, Iowa.

Mood Music by W. V. Bingham, director of the Department of Applied Psychology, Carnegie Institute of Technology.

Measures of Musical Talent by C. E. Seashore of Iowa State University.

Baldwin's Public School Music Test by Ralph L. Baldwin, 81 Tremont Street, Hartford, Connecticut.

Scale for Measuring Ability in Sight-Singing by E. K. Hillbrand, head of the Department of Education, Dakota Wesleyan University, Mitchell, South Dakota. Work was done at Northwestern University, University of Chicago, and Columbia University.

For the benefit of those who are interested in the field of educational research in music the writer offers the following suggestions for future research. The problems mentioned here have come to the attention of the writer during six years of work in this field.

Problems for Future Research

1. Construction and standardization of a silent-reading test in vocal music. The only attempts in this field are the silent-reading selections in the Gildersleeve test, the few selections in the Beach test and the few selections in the Detroit Practice Tests. This could be a group test. An oral reading test must of necessity be individual.

2. Construction and standardization of a separate test for intervals, introduced in the order of their difficulty for the pupil. The degree of difficulty of the intervals might be determined by the use of the writer's sight-singing tests.

3. Construction of a test involving "rhythmic patterns" of increasing degrees of difficulty. This suggestion was urged especially by Miss Curry of Columbia. This would be a test for rhythm instead of intervals.

4. Construction of a test with interlaced selections, one selection involving intervals and the other involving rhythmic patterns. This type of test could be used for diagnostic purposes.

5. Construction of a test with selections in minor mode interlaced with selections in major mode. Some light might be thrown on the problem of whether pupils sing selections written in minor keys as easily as those written in major keys.

6. Construction of a test where the single varying factor is the *key* in which the selection is written. Can the average child sing easier in one key than another?

7. Construction of a test with selections using *words*, interlaced with selections in which the *syllables* are used. What is the importance of the verbal element in sight-singing?

8. Construction of a sight-singing test using the intervals which occur most often in songs. These might be compiled by obtaining from 100 fourth-grade teachers, 100 fifth-grade teachers, and 100 sixth-grade teachers a list of five songs which their pupils can read at sight at the end of the school year.

9. Make a thorough investigation of eye-voice span in vocal music.

10. Make a thorough investigation of the number of eye-fixations, their order and their length of pause for the *different types* of music. There are, without doubt, marked differences between the reading of a simple one voice selection and a complex orchestral score.

E. K. HILLBRAND

Dakota Wesleyan University

Silent Reading in High School

Three investigations are now in progress under the direction of the New York Society for the Experimental Study of Education. One of these is the inquiry into silent reading conducted by Section 25, Secondary English. The inquiry is taking two directions. The first is to discover what relation subsists between a pupil's ability to read and his attainments in various fields of classroom endeavor. To determine his ability to get the thought from the printed page, three forms of the Thorndike-McCall Reading Scale have been employed. His intelligence has been measured by two forms of the Terman Group Test of Mental Ability or by a combination of the result on a Terman Group Test with his standing in the Otis Group Intelligence Scale. His achievement in three fields for two years—(1) English, civics, and history, (2) mathematics, and (3) foreign languages—is transcribed to cards from the school records. Most of this material for some two hundred and fifty pupils has now been gathered. The next step will be to work out the correlations.

The second direction in which the inquiry is moving is to compare the standing of a set of classes taught during some fifteen weeks with emphasis on getting the exact thought from the page with another set taught without unusual emphasis on thought. There will be an intelligence rating for each pupil in both sets of classes. It is impossible, however, in the New York City high schools to arrange these classes so that the teachers conducting the experiment shall have two classes of equal or equivalent ability. Much less would it be possible to pair pupils in the parallel classes. In one of the schools, Form 7 of the Thorndike-McCall Reading Scale was made the basis for rearranging the pupils according to ability. Two other forms were made the measure of the pupils' reading ability. In other schools, three forms of the scale were used to measure the pupils' ability, without any effort to regroup the pupils. In all schools, two additional forms of the scale were given at the end of the term to provide the final measure of reading ability. The assumption is that the parallel classes of the ten teachers involved will furnish

a statistical basis for determining the effectiveness of teaching with emphasis on the thought. The correlations with the intelligence rating of the pupils may indicate whether the difference in teaching produced the difference in the result. The investigations of Dr. Arthur I. Gates indicate, however, that the Thorndike-McCall Reading Scale measures chiefly depth of comprehension and that its scores parallel pretty closely the intelligence scores. The chances of securing very significant results with this part of the experiment are not so bright as with the first part.

These investigations have been started and will be carried through by means of a subvention provided by the Commonwealth Fund. The expenses for the materials and the clerical assistance, therefore, are to be met through the generosity of the Fund. A good many teachers, in several schools, have contributed their time and effort. The patience and zeal of these teachers who have been engaged in the experiment is a proof of their professional spirit. English teachers in New York struggle under a heavier burden than in many cities, so that any contribution they make toward educational research involves a genuine sacrifice on their part.

DUDLEY H. MILES

Chairman of the English Department,
Evander Childs High School, New York City

Educational Research Association

(FRANK N. FREEMAN, *Secretary and Editor*)

REPORT OF THE ANNUAL MEETING, CLEVELAND, 1923

The annual meeting of the Association was highly successful. At the first open meeting there was an attendance of 800 and great interest was shown in the papers which composed this program. The detailed program was printed in last month's issue, so it will not be repeated here.

The closed meeting was held Thursday forenoon. There was a small attendance at this meeting, but the significance of the session was not in any way indicated by the size of the audience. Since the program of this session was not printed last month, brief comment on it may be made. In the first paper Dr. Karl J. Holzinger of the University of Chicago presented a technical critique of statistical methods as applied to mental and educational tests. He gave new formulae for the determination of the reliability of tests and presented evidence concerning their scoring and organization. He held that in some respects tests should be organized and scored more simply than is the prevailing practice. The second paper, by Mr. E. R. Wood of the State Normal School of Emporia, Kansas, was entitled "A Simplified Process for Obtaining Partial Correlations and Regression Equations of Three or More Variables." Mr. Wood has worked out a graphical method which

makes it possible to calculate partial correlations and regression equations in a few minutes. The members present were much interested in this new method.

A somewhat similar problem was attacked in a different way by Mr. Courtis who gave a number of tables to illustrate the procedure which he was describing. His procedure consisted in selecting individuals for comparison who should be equal in the various aspects which might affect the measurement in question. The last paper, entitled "A Percental Graph," was presented by Dr. A. S. Otis in which the percental graph was used to illustrate the method of comparing scores in two different tests.

These papers were discussed as a whole by Dr. Leonard P. Ayres. The members of the Association were greatly pleased to have Dr. Ayres meet with them and offer his illuminating comments upon the problems which were being considered. Several other members of the group also contributed to the discussion.

The third program was held in a large auditorium on Thursday afternoon. At this meeting Dr. Ayres presided. Since this meeting was on the last day, the attendance was smaller than it otherwise would have been. The general subject of the program was "Educational Costs and the Building Program."

The annual dinner was held on Thursday evening. An enthusiastic group of 50 attended the dinner and listened to the comprehensive, well-balanced, and critical survey of educational investigations for the past year presented by Dr. Theisen, the retiring president.

A number of committees, which had been appointed at the preceding meeting, made reports. The first of these was the Committee on Educational Journals—chairman, Dr. Rugg. The committee reported that it had canvassed the relationships between a number of the educational journals in order to determine the extent of the overlapping in the subject matter of the articles printed in these journals, and considered further whether this overlapping was undesirable and might be avoided. It appeared from the study which was devoted to this matter that it was undesirable to attempt any formal or exact definition of the field of the respective journals. It was informally agreed that the JOURNAL OF EDUCATIONAL RESEARCH should emphasize articles of a practical import and that the *Journal of Educational Psychology* should be recognized as dealing especially with the more technical articles, but this arrangement is a purely informal one and binds neither journal to any definite action.

This committee also had referred to it a suggestion of the National Research Council that the society should consider the project of making an annual review of the literature of educational research. The committee considered that the association might profitably make an experiment in this direction. It learned that Dr. Buckingham, in connection with the N. E. A. committee for the coordination of educational research, had undertaken to collect a bibliography of educational research. Dr. Buckingham further offered to provide the Association with copies of this bibliography. The committee recommended that a committee be appointed by the president to undertake

this annual review of educational research literature for one year as an experiment and it should appoint whatever individuals it found desirable to be responsible for certain divisions of the field. This recommendation was seconded and the motion carried.

The second committee to report was the one on reorganization, and the following is the report of the committee—chairman, Dr. Woody.

"The committee appointed by the president of the Research Association to consider the advisability of reorganizing the Association recommends that the present form of organization be maintained. In making this recommendation the committee desires to say that it took into consideration the following facts:

1. That 92 percent of all members replying to the inquiry sent out, reported that the present form of organization is well-adapted for satisfactorily meeting the needs of its members.
2. That under the present organization there is great flexibility in the arrangement of programs.
3. That the interests of a majority of the members are general rather than highly specialized."

This report was adopted.

The Committee on Affiliation or relations with other organizations, chairman, Dr. Whipple, reported that it had held a meeting during the autumn of 1922. At this meeting the plans of the Association were canvassed in relation to the plans of the National Society for the Study of Education and the National Society of College Teachers of Education, both of which happened to be represented on the committee. It was found that the plans for the program of the three associations for the coming meeting did not conflict, either in the subject matter of the programs or in the time. It appears that the three associations are spontaneously developing differentiated functions. This being the case, the committee recommended that the program committee of the association keep in mind the desirability of avoiding much overlapping and of communicating with the other associations at the time it is making its plans for the annual meeting, in order that such overlapping might be avoided. This report was adopted.

The Executive Committee considered an application from an organization for educational research on the Pacific Coast for affiliation with the Educational Research Association. This matter was presented by Dr. Dickson. The Executive Committee appointed a sub-committee consisting of M. E. Haggerty, V. E. Dickson, and Frank N. Freeman to draw up amendments which should provide for such affiliation and present them for consideration. These amendments were drawn up and approved by the Executive Committee. They were then presented to the Association. The following were the proposed amendments:

1. Regional organizations devoted to educational research may become affiliate members of this association upon recommendation of the Executive committee and approval of the association.
2. Only organizations whose purpose and activities are in harmony with this association may become affiliate members, and in each case there shall be a reasonable assurance that such a regional organization gives promise of more than temporary existence.

3. Such affiliated societies may elect a representative to this association who shall be a member in good standing of this association.
4. The representatives of affiliated organizations together with the members of the executive committee shall constitute a council for the purpose of considering the problems arising out of affiliation.

The proposal aroused lively discussion. It was generally recognized that it was desirable to encourage educational research and to promote sympathy with and an understanding of research among those who might not be qualified for membership in the association. Although some fears were expressed that a formal affiliation would not serve any definite purpose, the association voted by a two-thirds majority to adopt the amendment. It was further voted to authorize the president to appoint a committee to report at the next meeting concerning the nature of the relationships of the society to affiliated organizations and the policy which should be accepted. A further motion directed that the executive committee should constitute this committee. The application of the Pacific coast organization was therefore postponed for consideration at the next year's meeting.

An amendment was adopted which makes the term of office of the Secretary three years.

The following officers were elected:

President—Clifford Woody
Vice-President—J. L. Stenquist
Secretary-Treasurer—H. A. Greene

The financial statement of the society for the year is as follows:

Cash on hand, April 24, 1922		$174.50
Receipts—From dues		537.90
Total		$712.40
Expenditures—		
Subscriptions to Public School Publishing Co.	$335.00	
Expenses of Executive Committee	68.00	
Expenses for clerical help, postage and printing	52.11	
Rebate on subscription overpaid	1.00	
	456.11	
Balance on hand	256.29	
Total		$712.40
Bills payable—		
Printing, postage, etc.	$ 38.18	
Net assets	218.11	

FRANK N. FREEMAN,
Retiring Secretary-Treasurer.

Audited—
P. R. Stevenson.
A. S. Otis.